Inference and the Foundations of logic

James R. Shaw

Draft of December 7, 2023

Contents

Preface

Note to Readers

Thanks to anyone taking a look at the circulating draft out of curiosity. I welcome feedback of any kind.

This book spans two separate literatures that often don't engage with each other: one in epistemology and the metaphysics of mind on the nature of deductive inference and its norms, and another in the philosophical foundations of deductive logic.

Though this book argues that these fields are deeply intertwined, I am hopeful that there are worthwhile components of the book even for those interested exclusively in one field or the other. For the reader merely interested in the nature of inference, a natural path through the book would be to read Chapter 2 followed by Chapters 4 through 6 (with an optional detour through Chapter 3 for those who want to delve more into the epistemic evaluation of inference). These chapters build on each other and should be read in order.

Those merely interested in the foundations of logic would probably be most interested in Chapters 2 and 7 (again with an optional detour through Chapter 3 for those specifically interested in the normativity of logic), followed by any of the applications explored in Chapters 8 through 11 of Part II. These latter applications can for the most part be read individually and out of order, though I should warn that all of them make periodic use of foundational claims from Part I. I'm hopeful that some readers may be able to get by 'black boxing' some of these appeals. But those who want to appreciate the full justification for the given approaches to logical problems would minimally need the resources of Chapters 2–5.

CHAPTER 1

Introduction

> "Judgments, in which one is conscious of other judgments as justifying reasons, are called *inferences*. There are laws governing this kind of justification, and to set up these laws of correct inference is the aim of logic." – Gottlob Frege[1]

This book is about deductive inference, insofar as it is illuminated by logical inquiry; and about logic, insofar as it is illuminated by the nature of deductive inference.

As my Fregean epigraph evinces, the idea that logic studies the correctness of a distinctive mental act of inference is not new. It has roots in Aristotle and finds expression (in one form or another) from virtually every canonical contributor to the field. Even so, this understanding of logic is controversial. For example, it stands in contrast to an equally prominent tradition taking logic to study general, or content-neutral, truths. And it is often conflated with closely related claims, such as that logic studies good reasoning.

Equally frustrating is that even if one accepts that logic studies inference, our understanding of inference in the analytic tradition is still incipient. This is easily seen within, say, the philosophical corpus of the 20th century, by comparing the meager portion devoted to inference with the sprawling body of work on mental states like belief or knowledge. The philosophical landscape is just now changing, with a spate of recent work taking up unresolved questions about inference. I have hope that this work has brought us to a point where links between logic and inference that were once liable to obfuscate are now instead capable of generating mutual illumination.

[1] Frege (1879?/1983, 3), emphasis in original, my translation.

I am tempted to say something even stronger: that we cannot fully understand the foundations of logic independently of the mental process of inference, nor can we fully understand inference independently of puzzles naturally expressed in logical terms. At some point, we must treat these two areas of inquiry in tandem. As a reflection of this conviction, this book systematically winds back and forth between discussions of the foundations of logic and of the nature of deductive inference, stepwise using insights from one topic to clarify our understanding of the other.

Before saying much more, however, I must straight away sound a methodological note of caution concerning my use of the word "logic". Debates over the nature of logic are ripe for verbal disputes. "Logic" is after all a term of art already plainly used to cover several conceptually distinct branches of inquiry. It seems advisable for those working in the foundations of logic to begin by saying what exactly they mean by "logic" or to try to demarcate their topic by some other means in order to avoid pointless debates. I will eventually offer a characterization of logic in non-logical terms, and given this I find it most helpful to frame my talk of logic as having both weaker and stronger possible construals.

According to the weaker construal, one can think of my eventual characterization of logic as stipulative: I aim to explore particular kinds of formal techniques for investigating the conditions on good deductive inference (where the sense of this claim unfolds gradually over the course of the following chapters). One could substitute the term "logic" with my eventual characterization at many junctures in the book and, it is my hope, retain a good deal of the work's interest. This is because the formal study of the conditions on performing a mental activity of deductive inference well should be a topic of independent interest, regardless of whether that study conforms well to *any* preexisting use of the term "logic".

That said, words can matter, and it is not for nothing that I invoke the term "logic" for my investigation. There is a core tradition in symbolic logic, especially as it is taught and applied by philosophers, with a rich and storied history. This tradition has increasingly splintered into debates among theorists clearly aiming to be responsive to a unified phenomenon of some kind. This is witnessed in the justifications adduced in (sometimes heated) debates over the tenability of certain logical inference rules. And it is witnessed in generalizations about logical subject matter used to attack, or defend, logical frameworks

like modal, epistemic, or higher-order logics.

According to a stronger construal, my use of "logic" is meant to capture a rational reconstruction of a central, often implicit or latent, preoccupation of this aforementioned core logical tradition with inferential goodness. My claim here is not that a preoccupation with inferential goodness matches the *self-conception* of many, or most, or even some privileged group of logicians. It is rather that core logical tradition is naturally and fruitfully seen through the lens of the characterization I develop.

The evidence for the value of seeing core logical tradition as preoccupied with inferential goodness is meant to accumulate gradually over the course of the whole book, and can be brought out with the following hypothetical. Suppose for the moment an investigator had never seen or heard of logic before, but could identify the mental activity of deductive inference roughly by ostension, and sought to better understand the conditions under which this activity was performed well. What I will be arguing is that this character would end up developing frameworks, formal-systems, concepts, and distinctions that mirror those from traditional logical modes of inquiry. They would, for example, naturally be drawn to rebuild classical logic to capture good inference patterns in semantically well-behaved domains like mathematics, and could equally be drawn to the use of model-theoretic techniques to model it. And they would naturally refine this framework to model inferences for modalized thought and discourse, to capture the possibility of semantic defect, to integrate perspectival information, and to reflect information-state-sensitive language in ways that would look strikingly like the ways logics have been refined for roughly those purposes over the past century.

What is more, because of the distinctive inferential foundations of this character's inquiry, I will argue, the resulting formal frameworks would be imbued with various kinds of significance that have influentially been claimed to belong to logic by distinguished members of core tradition. These would include the claims that logic is relevant to reasoning, that it is prescriptive or normative, that it gives something like constraints on intelligible thought, that logical truths have a privileged metaphysical and epistemic status, and so on. The inferential conception thus illuminates what logicians have said about logic and what shape various logics have taken, sometimes by simply vindicating logicians' pronouncements, and sometimes by providing us with tools to diagnose tempting but ultimately problematic assumptions that made the

pronouncements seem reasonable.

The upshot at the end would be this: that the inferential conception gives *one* branch of inquiry that is as worthy as any other of bearing the heavy historical mantle of the philosopher's "logic". I will not directly argue in this work that there are not other branches of inquiry which may meet this high standard as well (though I will present some noteworthy concerns for rival conceptions as I go). Still, even situating the inferential conception of logic among several best candidates to match the tradition would be significant. Numerous theorists have made general pronouncements about logic, clearly invoking core tradition, that conflict with the inferential conception (i.e. they have made pronouncements that logic is *not* normative, *not* relevant to reasoning, that logical truths do *not* have a privileged epistemic or metaphysical status, etc.). This is typically done without in any way qualifying their use of the term "logic". Such unqualified claims would at best be misleadingly ambiguous, and at worst false. The claims could perhaps be rescued by explicitly qualifying them to accord with some specific, stipulated branch of inquiry (ideally in non-logical terms) other than the inferential conception that was argued to accord well with core tradition on independent grounds. The significance of the general claims, suitably restricted, could be worth thinking through, but certainly would be substantially diminished.

I hope to have made a good case for the stronger construal of my claims about "logic", and will return in the book's conclusion to review some of the evidence for thinking this. Still, I recognize that some readers may not make it that far, and that others who do may not be persuaded. For these readers I want to emphasize the weaker construal is always available: drop any preoccupation with the word "logic" and remain open to the inherent richness of a formal investigation of a mental act of deductive inference for its own sake. There could still be much of value to take away.

Having flagged this distinction between weaker and stronger readings, I will now proceed to speak as if there is just one thing at issue—logic—without further qualification, and to lean into the stronger ambitions of the book. And with the book's overarching aims in this regard having been previewed, let me describe how I try to achieve them.

My simple point of departure is the assumption—buoyed by longstanding philosophical tradition—that there is a distinctive mental activity or event of inference with a broadly familiar shape. Chapter 2 collects features commonly

attributed to inference, and deductive inference in particular, while developing two more controversial lines of thought about it.

The first line of thought is that we can read important features of inference off of the distinctive structure of the mental states it mediates between. These are the information-bearing mental states—like belief, supposition, and imagination—which characteristically bear truth-conditional structure (as opposed to preferential or inquisitive structure). I exploit inference's exclusive ties to such states to anchor a role for truth in the structure of inference. The second controversial line of thought takes the fact that 'goodness' is attributed to inference to establish its status as a goodness-fixing kind, and explores links between that standard and truth. The result of integrating the two foregoing lines of thought is a familiar (though again, controversial) view: that good deductive inference necessarily preserves truth for some modality. The grounding of this familiar thesis in a more fundamental theory of inference and inferential goodness is intended to give it a fresh appeal, along with some new justification that guides its application.

With this, still skeletal, conception of deductive inference on the table, I propose to think of logic as a certain highly constrained way of investigating good deductive inference. At this early juncture it is probably best to bear in mind the possibility of taking this proposal stipulatively, with the justification for the utility of the stipulation to unfold. Even seen as provisional stipulation we can ask: What should logic look like *so-conceived*? What importance would it have?

This leads to the first way a theory of inference can inform the foundations of logic in Chapter 3: by opening up new space to understand the normative significance of logical techniques and their relevance to reasoning. I argue that if we construe logic as the study of good inference two things follow. First, any normative implication of logic should be understood in *evaluative* terms (like 'goodness'). Second, the relevant evaluative notions would govern a mental *act, event, or process*. I note that the recent literature on logic's normativity is systematically marked by either the presupposition that it will be cashed out in terms of what we epistemically *ought* or have *reason* to do, or the presupposition that it will govern mental states (like belief or credence). I explain why trying to capture *any* evaluative notion governing acts (e.g. the goodness of a good fastball pitch) in terms of what states one ought to be in leads quite generally to a series of predictable obstacles. I then note that these are precisely

the kinds of obstacles we find hampering and constraining current attempts to cash out logic's normativity. Once we free ourselves of the presuppositions of the debate, a simple, exceptionless principle capturing logic's normative force can be formulated that skirts this array of obstacles. And in the process, we can also uncover a constrained, but distinctive, role for logic in the study of reasoning more broadly.

These lessons for the relevance of logic can be appreciated even with the mere skeletal conception of good inference that I begin with in Chapter 2. From there, I turn to a puzzle that draws on issues historically framed in logical terms, whose resolution can be used to flesh out that skeletal conception of inference. The motivating idea, which finds expression in Aristotle, Aquinas, Descartes, Spinoza, Hume, Kant, Husserl, Wittgenstein, and a small ongoing tradition that takes up their ideas, is that certain logical impossibilities seem to resist thought. What is most perplexing is that these impossibilities appear to resist not merely belief, but also supposition or imagination. I bolster this claim with some modest experimental work. In the process, I temper the claim with evidence that increasing the complexity of a logical impossibility relaxes cognitive resistance, making a given impossibility easier to entertain. Putting these elements together, I argue in Chapter 4 that we best explain all these phenomena by positing a special and demanding cognitive relation between an agent and a proposition that I call (*representational*) *crowding-out*, which precludes the agent from representing that proposition. I explain how this relation arises organically on 'mirroring' conceptions in the foundations of mental representation. I then show how those conceptions give us important tools both to understand how impossibilities obstruct thought as well as how complexity can work to relax that obstruction.

I then claim in Chapter 5 that the crowding-out relation is the key element lacking from our current understanding of deductive inference. Virtually every commenter on inference has noted that an agent consciously performing an inference somehow 'takes' or 'appreciates' their inference to be a good one as it is performed. The varied and often incompatible proposals for what constitutes this 'taking'—a belief, an intuition, a form of rule-following, an 'inferential force'—are testaments to the difficulties in capturing its unique features. I argue that the cognitive relation of crowding-out turns out to be ideally suited to explain a host of unusual features attributable to deductive inference. On the resulting view, deductive inference is a way of settling a question in inquiry

by reducing the space of thought so that only one answer to a question is thinkable. Using this idea, I formulate a reductive analysis of deductive inference in terms of crowding-out representations, and show that this analysis skirts numerous problems in understanding deductive inference (e.g., worries from deviant causal chains and Carrollian regress) while also illuminating some of its most telling features (e.g., that inference tends to proceed in relatively small steps, and that trusting reliable testimony about an inference's goodness does not generally position one to perform the inference). I also show the account naturally integrates into a broader, unified account subsuming ampliative inference in Chapter 6.

So with Chapter 3 we see how a skeletal conception of inference can transform our understanding of the goal of logical inquiry. And with Chapters 4–6 we see how reflection on logic's relationship to possible thought can help transform our skeletal picture of inference into a full-fledged analysis. With Chapter 7, I wind back again to apply the newest lessons about inference to logical theorizing. By supplying a reductive analysis of deductive inference alongside a characterization of logic in terms of inference, I claim we have the tools for a reductive analysis of logical consequence relations themselves. Using the work of John Etchemendy as a foil, I argue that these analyses have the potential to do precisely the work we would hope of them: that of reducing vexing logical problems to questions in non-logical domains whose resolution does not prejudge relevant logical matters.

To further justify this claim, I give a series of applications of my analyses to various logical frameworks. I try to show both how the structure and limits of the relevant frameworks are illuminated by the analyses, but also how the analyses reduce questions about particular contested logical rules to questions in the philosophy of mind, the philosophy of language, linguistics, and metaphysics.

Chapter 7 begins this process. On the conception of logic I favor (which I should flag has many noteworthy antecedents), logic investigates necessary truth preservation among the contents expressed by sentences in virtue of a subset of their linguistic properties, as a means of gaining clarity on the conditions on good deductive inference. I note that if we are careful about the selection of a class of linguistic properties possessed by sentences of a first-order language, we retrieve ordinary first-order logic as a result. This safeguards the importance of classical logic for domains of discourse which possess the lin-

guistic properties in question—domains that are 'semantically well-behaved' in certain respects. Mathematics provides a key instance of such discourse. But the result also reveals limits to the application of first-order machinery to discourses with less well-behaved semantics (even if those discourses are superficially 'regimentable' in first-order form). From there, I turn to consider the contested classical rule of Ex Falso, and the contested classical principle of Excluded Middle. I explain why Ex Falso should be safeguarded in our theories (even though it hardly, if ever, represents 'good reasoning'), and clarify what kinds of foundational and empirical claims in the philosophy of language and linguistics would need to hold for a domain of inquiry to overturn Excluded Middle.

Chapter 8 turns to a small debate over how to define validity for logics that contain two-dimensional modal operators that appears to put the idea that logical truths are metaphysically necessary in jeopardy. Leaning on features of my framework for logic, I argue that the terms of this debate are partly based on a conflation of semantic value and assertoric content. Once the relevant distinctions are drawn, I claim that two apparently rival definitions of validity do not stand in competition, but can represent different stipulative choices for the kinds of assertoric contents expressed by sentences of a language containing modal operators. And on either stipulation, logical truths would remain metaphysically necessary.

Even so, I acknowledge that parties to the debate over the correct modal logic were sensitive to some genuinely puzzling phenomena. I use the contested validities of the debate to raise a new puzzle about the nature of deductive inference that cannot be resolved by stipulation, precisely because of its ties to ordinary linguistic usage. Roughly, the puzzle is that certain inferences appear to be good when performed on the basis of believed premises, but no longer appear to be good when they are performed on the basis of premises that are counterfactually supposed. This fact applies pressure to the otherwise plausible thesis that the goodness of a deductive inference depends purely on the contents of its premises and conclusion. I conclude the chapter by framing my preferred resolution of this puzzle. But I flag that defending this resolution leans on highly contested theses in linguistics and the philosophy of language. I take the discussion to reveal some limited but important ways in which empirical matters could indirectly bear not only on certain logical inference rules, but on the foundations of logic itself.

Chapter 9 turns to examine how logics should be adjusted if they accommodate the presence of strong semantic defect: the failure of truth-evaluability either due to a sentence's failing to express a proposition, or to a sentence expressing a trivalent proposition. I argue for two lessons in this setting. The first lesson is that although the presence of defect in a system tends to lead to logical 'weakness,' in the sense of licensing fewer entailments, this apparent weakening is better understood as a process of rebranding some instances of logical entailment as instead being instances of general entailment. In this way, logics of defect rarely treat formerly recognized good inferences as bad inferences, instead of as good inferences of a non-logical kind. The second lesson concerns how validity and consequence are relativized to a stipulated set of 'logical' linguistic properties. I discuss how the presence of semantic defect creates a theoretical choice-point owing to this relativization that leads to two different, but equally legitimate consequence relations for defect. Perhaps surprisingly, one of these is simply classical logic.

I apply both of the foregoing lessons to logics developed for formal theories of truth that treat liar-like sentences as defective. In particular, I consider a style of objection frequently raised against such theories that they give rise to a logic that is 'too weak' to carry out ordinary reasoning. I argue that these objections run afoul of both lessons from earlier in the chapter, especially when directed against theorists who are suitably upfront about the character of the defect that is perturbing logical relations within their system. Saul Kripke will provide a clear example of this kind of theorist.

Chapter 10 explores how logics should be reframed in response to three interrelated phenomena: perspectival thought, context-sensitivity, and ambiguity. I begin by reviewing Kaplan's seminal logic LD for perspectival context-sensitive terms like "I", "now", and "that". I highlight a curious existence entailment ("I exist") within Kaplan's system, and discuss an apparent challenge the system presents for thinking that logical truths are metaphysically necessary. I use these topics to motivate an independent investigation into how a logic should be adjusted to accommodate perspectival or '*de se*' thought—noting how questions about such a logic can be raised and addressed independently of questions about language (including linguistic context-sensitivity). I develop a logic for this setting, LD^*. I note that while LD^* is a minor variant of Kaplan's LD, the former nevertheless invalidates LD's controversial existence entailment and has very different philosophical underpinnings that

conceptually connect validity to *de se* modality—essentially a generalization of metaphysical necessity.

From there, for instrumental and illustrative purposes, I explore how a logic should adjust in response to the presence of lexical ambiguities, arguing that logics which forgo resources to resolve ambiguities under-generate in predictable ways and become incapable of describing good deductive inference in the manner that other logics can. I then lean on this investigation to explain why one would expect some parallels between a treatment of linguistic context-sensitivity and ambiguity. Following this idea up, I argue that several of Kaplan's modeling choices in developing LD are far from obligatory (and even in some ways idiosyncratic) in the task of developing an inferential logic even for the perspectival context-sensitive terms that Kaplan made his focus. I then note these problems are exacerbated with respect to 'non-perspectival' context-sensitive terms (including gradable adjectives, quantifiers, and modals). Drawing on all the points within the chapter, I conclude with a conjecture: that Kaplan's logic for context-sensitive terms reflects a periodic conflation of linguistic context-sensitivity and perspectival thought with the result that Kaplan's system, and the philosophical basis underlying it, are a kind of hybrid that fails to faithfully model either phenomenon.

Chapter 11 discusses the question of how to develop logics for languages whose semantics make use of a shiftable information-state parameter, typically in application to epistemic modals, conditionals, or other expressions bearing the hallmarks of the language of subjective uncertainty. I trace out several lines of thought leading to information-state semantics beginning with Vann McGee's putative counterexamples to Modus Ponens. All these lines of thought seem to press the question of what the 'correct logic' for information-state semantics should be.

I argue that the answer to this question is heavily dependent on the broader framework in which an information-state compositional semantics is interpreted and applied in ways that are not often acknowledged. I try to justify this claim by looking at two systems, given by Seth Yalcin and John MacFarlane respectively, which employ very similar compositional treatments of conditionals and modals but, I argue, should give rise to strikingly different applications of logical machinery in the context of modeling deductive inference. In particular I argue that the most perspicuous logic for Yalcin's framework is given by its modal- and conditional-free fragment (with a result that could

be as simple as classical logic). MacFarlane's logic by contrast must integrate modal and conditional language, though its details cannot be fully ascertained due to a subtle circularity arising within MacFarlane's account of the information contained in mental states.

From there I turn to explore two popular conceptions of validity for information-state logics sometimes called 'classical' (or 'diagonal') consequence and 'informational' consequence. I note that there is a tendency to conflate a rejection of the former consequence relation with a rejection of logic as tracking relations of truth-preservation. Focusing on the work of Justin Bledin, I argue that this tendency arises from a conceptual confusion. Once the typical application of information-state semantics is taken into account, we see that informational consequence is the most natural extension of the view that logic is concerned with the necessary preservation of truth (though this fact is admittedly obscured by its typical formulation).

After a brief return to explore how McGee's counterexamples to Modus Ponens interact with a recent literature on the 'weakness' of belief, I conclude by discussing a trend in information-state semantics to treat probability modals using probabilistically graded attitude states like credences. I remind that we currently have no adequate models of how to reason, let alone infer, with graded mental states like credences and that this interferes with our ability to give any sense to a deductive inferential logic in this context. I do my best to make some first steps in developing a framework for inferring with graded attitudes. But I note that this attempt requires us to stray far from the standard motivations and foundations for graded mental states, and opens the account to numerous foundational obstacles, each of which could undermine the intelligibility of a distinctive logic for deduction in the context of graded mentality.

It is important to preview that Chapters 7–11 leave *many* questions outstanding. Often these questions have very far-reaching implications. For example, we will run up against questions about the correct theory of natural language assertoric content in Chapter 8, the relationship between linguistic semantic defect and the structure of mental content in Chapter 9, the nature of *de se* cognition in Chapter 10, and the question of whether mentality is fundamentally structured in truth-conditional or probabilistic terms in Chapter 11. Addressing each of these issues in full detail would call for a separate book-length treatment, and the outcome could radically reshape our understanding of deductive inference, and so also bear on the very tenability of my conception

of logic. Other times, and just as often, the questions left open in these chapters concern much more minor questions about the validity of some particular inference rule. I regard all these open questions, large and small, as inevitable, if not welcome in the context of my limited investigation. The goal of giving my account of logical consequence is not to supply a clean resolution of several tough questions in philosophical logic. Rather, it is to provide some tools that could allow for definite progress on those questions, while hopefully respecting their difficulty. Let me elaborate on that point just a little, since it constitutes another of the book's main ambitions.

It has long been a point of dissatisfaction just how rapidly logical debate can turn into brute intuition mongering. Not only is there a sense that logical intuitions are justificatorily shallow, but there are obvious reasons to think that such intuitions are highly sensitive to indoctrination and other forms of bias. Philosophers have rightly started to probe deeper into logic's foundations to look, not necessarily for answers, but for spheres of inquiry in which persuasive answers to logical questions could gradually be developed without simply prejudging the original questions. A recent influential pursuit of this kind is witnessed in the recent outpouring of work on the normativity of logic. A guiding thought of that literature, expressed by John MacFarlane, is that if logic somehow governs good reasoning, perhaps we could use what are hopefully less controversial intuitions about good and bad reasoning to work backwards and resolve questions about more controversial logical principles.

As I discuss in Chapter 3, I think MacFarlane's particular strategy for adjudicating logical principles ultimately does not pan out. Logic's ties to good reasoning are indirect (on any reasonable conception of "logic"). And to the extent that logic bears on good reasoning, judgments about the latter will be swayed by the very same biases that inform judgements on the former. We really get no deeper, justificatorily, by examining good reasoning. In spite of this, I think that MacFarlane's broader idea—that to make progress on logical questions we need to get much clearer about what phenomena logic is supposed to be modeling—was exactly the right one.

As recently flagged, one thing I hope to be exemplified in this work is a defense of the fruitfulness of investigating alternative inferential foundations for logic. The shape that various logics have historically taken fits extremely well with the inferential conception. The foundations of logic are illuminated by seeing them in inferential terms. Debates over logical principles are clarified.

And new routes for resolving logical disputes in non-logical terms abound. There is still much work to do both in justifying and applying the framework, but I hope to have made a respectable start.

And even if my proposed conception of logic falter, I hope it can still lend some conviction to the idea that the foundations for logic are potentially varied and complex. If so, logical disputes would not be arbitrated on the basis of brute intuitions about logical relations, nor on the basis of oversimplified metrics like simplicity or theoretical strength, nor even on the basis of cursory claims about logic's ties to reasoning, truth, inference, or any other concept. Instead, logic could draw on diverse and substantive claims scattered throughout disciplines like the philosophy of mind, the philosophy of language and linguistics, metaphysics, and epistemology. In short, my hope and my aim is to make the case that there are opportunities to locate rich nonlogical foundations for logic. Probing such foundations could embroil the logician in myriad forms of nonlogical controversy. But it also could, precisely by tying logic to non-logical domains, lay the groundwork for the possibility of lasting foundational logical progress.

.

Part I

Foundations

CHAPTER 2

Inference, Inferential Goodness, and Logic: a Skeletal Account

Let me begin by giving an account of inference and inferential goodness, and then say how we can understand logic as investigating the latter. This account be will schematic as my provisional goal is merely to get those features of inference on the table necessary to understand its relationship to logic and—especially for the next chapter—to clarify how that relationship could imbue logic with some normative significance.

First, we need to disambiguate "inference". This word can be used to denote something abstract—for example, an ordered pair of propositions in which the first serves a premise and the second as conclusion. It can be used to designate a transition within a particular formal system—for example, for a rule like Modus Ponens as used in the context of a natural deduction system. It can also be used for cognitive processes that manage information in-principle unavailable to cognizing agents—for example, for manipulations of information about shapes and distances in human visual systems.

I will not be discussing any of these senses of "inference". The use of the word that interests me denotes a simple, familiar person-level *mental process* or *event*. For example, I believe that Joppa is north or south, and believe it's not north. From these I infer that Joppa is south. Here is a sense of "infer" that applies to a rationally evaluable mental act or event. The inference is a mental happening of sorts, and one of which we can typically be consciously aware (even if we are not always so aware). And it is not merely that the starting belief states 'bring to mind' the concluding state. Rather, the latter state is somehow 'drawn' from them in a way that is subject to rational evaluation.

At this point, I do not expect this characterization of inference to be en-

tirely clear. What is hopefully clear enough is that there is *some* phenomenon to be investigated that does not involve mere abstractions, formal systems, or general cognitive information processing. What is more, this notion of "inference" should not (or at least not without substantial argument) be identified with *reasoning* broadly construed to include processes by which our beliefs and other attitudes are rationally revised. For example, reasoning in this broader sense can lead to the abandonment of attitudes. But it is not clear one can 'draw' the *absence* of an attitude 'on the basis' of pre-existing ones in the same sense as my concluding belief above. Of course, inference appears to be a part of reasoning in the broad sense of 'reasoned change in attitudes.' But we should not presume it constitutes the whole of it.

I should also flag that person-level inferences will occupy center stage in this work. These are inferences of which we are, or can easily become, aware merely by making them. This focus is simply because these forms of inference are those we can learn about most easily through philosophical techniques of introspection and reflection. I do not mean by this focus, or by anything else I say in this work, to prejudge the question of whether there can be subpersonal or unconscious processes of inference.

Thus I aim essentially by ostension to get a rough grip on the phenomenon of inference we are out to investigate. Now: what is an inference in this sense? In asking this question, I want to propose a necessary condition to help to get us started: inference is a process in which one *acceptance state* is generated on the basis of others.

In the inference I gave above, I came to a new belief on the basis of two preexisting beliefs. But, crucially, I could equally *suppose* that Joppa was north or south, suppose further it was not north, and infer 'under supposition' that it was south. In so doing, I merely suppose something new on the basis of two suppositions. Likewise I seem able to infer my conclusion from imagined premises, if imagination is different from supposition.

Belief, supposition, and imagination are what Stalnaker (1984) calls states of *acceptance*, offering as a diagnostic that these are states we call 'correct' just in case their contents are true. One believes, supposes, or imagines correctly just in case one believes, supposes, or imagines what is the case.[1] Acceptance states contrast with preferential states like desire or hope, which are

[1] I will not be presuming, nor do I think, that the use of "correct" here tracks anything interestingly normative.

not said to be correct if what is preferred transpires. And they contrast with inquisitive states (like a state of wondering whether Joppa is north) which do not even take a truth-evaluable object. Notably, these latter states cannot participate in inference: one cannot infer from or to a state of desire or hope or wonder.

A different, but equally important understanding of acceptance states is as *information bearing* states—those in the business of representing how the world is or could be. The intentional structure of total mental states of this kind are fruitfully modeled by something like sets of worlds, or sets of propositions. Not so for preferential states which represent how a world is preferred to be, and are familiarly modeled not as a collection of worlds or propositions, but as a *ranking* of worlds or propositions.[2] Not so for an inquisitive state, like a state of wondering, which is better modeled by something like a *partition* on worlds, or sets of sets of alternative propositions.[3]

What this tells us is that inference is an act bridging all and only mental states with an inherently information-bearing, correctness-governed character. This helpfully constrains our understanding of inference if, as I will propose, we should aim to understand what inference *is* along broadly functionalist lines by what it *does*. We should construe inference as a mental activity which has some role to play in application to information bearing states and *only* to such states.

To this end, I now offer just such a broadly functionalist characterization of inference. A good part this proposal is unoriginal and much of it may seem platitudinous. I will elaborate on the choice-points in my characterization and note their historical antecedents. But I will not seek any extensive justification for these choices here. Rather, the hope is that their justification will lie in the fruitfulness of the characterization in its applications throughout the course of this book.

With this in mind, my working proposal is as follows:

> *An inference is a mental act or event whose proper function is to appreciably generate new acceptance states on the basis of old ones in a reliably correctness-preserving way.*

Let me unpack some of the terminology.

[2] See Bolinger (1968), Stalnaker (1984) Farkas (1985), Heim (1992).

[3] See Friedman (2013a) on inquisitive attitudes and Groenendijk & Stokhof (1997) for a discussion of the supporting literature on the semantics of interrogatives.

…a mental act or event…: Following prevalent views, I take up the idea that inference is not a particular kind of mental state (like belief), but an act or activity that mediates between such states.[4] But although I will sometimes speak of inference as an act or activity, not much hinges on this construal. If one is skeptical about whether inference is an act,[5] the ensuing claims about inference's relations to logic can all plausibly be defended on the weaker assumption that inference is a mere process or event. What is critical is that *inference is not a state*.

…proper function…: In speaking of inference's function, I appeal to a broadly functionalist tradition in the philosophy of mind according to which mental phenomena are individuated by their functional role in a cognitive system. More specifically, in speaking of *proper* functions, I appeal to a version of functionalism incorporating teleological elements, which is how a form of normativity will enter the picture. On this functionalist tradition, we conceptualize certain entities, including biological entities like hearts or kidneys as being *for* certain ends or purposes, like pumping or filtering blood. In being conceptualized with purposes, such entities establish a standard by which they can be said to be good instances of their kind.[6] For example, a heart is good (*qua* heart) provided it pumps blood well, bad (*qua* heart) if it pumps poorly. It is indisputable that inferences can be classified as good or bad. I propose that when we call an inference good we draw on a functional conceptualization of inference, associating it with a purpose establishing its attendant standard of goodness.[7] That purpose is spelled out by the three further features explained below.

…correctness-preserving…: Correctness for an acceptance state is, we have noted, truth. The conditions that matter to the correctness of such

[4] See Buckareff (2005), Gibbons (2009), Hieronymi (2009), Mele (2009), Peacocke (2008).

[5] E.g. see Strawson (2003), Setiya (2013).

[6] See Ziff (1960), Finlay (2004, 2014), Thomson (2008). The broad set of ideas here of course goes back at least as far as Aristotle. The application of them to the psychological sphere has its earliest detailed development in Millikan (1984), though I don't mean to plump for any particular (for example evolutionary) understanding of these teleological conceptions here—a general framework neutral on these issues will suffice.

[7] Cf. McHugh & Way (2018) who claim that *reasoning* (broadly construed) is a goodness-fixing kind. While I'm broadly sympathetic with this claim, I take no stand on it in this work, much less on what good reasoning generally requires.

a state are, therefore, those that matter to truth. These, familiarly, are encapsulated in the notion of a possible world. So correctness is world-relative. Say that an acceptance state transition is *correctness-preserving* at a world w just in case the based attitudes are correct at w if the basing attitudes are. The current proposal is that inference's proper function is to effect a correctness-preserving operation of some kind. One key piece of support for this proposal is that it has the virtue of *explaining why all and only acceptance states can participate in inference*: only they have the correctness-governed character which inference works to preserve.

…appreciably…: A conscious, successful inference is one whose distinctive correctness-preservation is somehow recognized by the inferrer.[8] When one consciously infers, one somehow 'takes' one's inference to be well-performed. Though it is notoriously difficult to specify what this appreciation comes to,[9] there are good reasons to consider the appreciation condition to be a rational commitment of the inferrer.[10] Though I do presume appreciability is a component of inference, I will remain neutral on what it involves until Chapter 5.

…reliably…: We've seen that correctness (or truth) preservation is world-relative. So which worlds matter to inference well performed?[11] A tempting option should be summarily dismissed: inference's function sometimes requires more, and sometimes less, than *actual* correctness-preservation. It may require more, because all inferences that start from suppositions of actual falsehoods preserve actual correctness of contents trivially. But obviously not all such inferences are well-performed. And

[8] Cf. Locke (1690/1979), Frege (1879?/1983), Russell (1920/1988), Stroud (1979), Thomson (1965), Sainsbury (2002), Field (2009), and Boghossian (2014). As noted before, I wish to allow for the existence of unconscious inferences. These needn't be appreciat*ed*, but seemingly need to be appreci*able* (to the inferrer) to play their rational role—hence the condition as stated.

[9] See especially the debate ignited by the exchange in Boghossian (2014), Wright (2014), Broome (2014), and Hlobil (2014).

[10] As argued by Hlobil (2019), consciously drawing an inference while believing it not to be good leads to a kind of Moorean rational incoherence. See the related form of Moorean incoherence discussed in Marcus (2012).

[11] Also which *kind* of worlds matter? Eventually, in Chapter 5, I will argue that good deductive inference requires correctness preservation at (all and only) metaphysical possibilities. But this claim, and the arguments for it, will not be needed for the intermediate conclusions of Chapters 3 or 4.

> it may require less, because ampliative inference may be well-performed even if it does not preserve actual correctness (because the actual world is an unusual or unlikely case). I think counterfactual inference helps reveal most clearly what inference, *qua* inference, aims to effect: it aims to preserve correctness across a 'safe' range of cases *compatible with the information contained in the starting acceptance states* on which inference operates. Ampliative inference may reliably preserve correctness in this way, even if it leads from an actually correct state to an actually incorrect state (since actuality may not have been among the 'safe' cases). And for suppositional inference to be reliably correctness-preserving may require more than being actually correctness-preserving, since many worlds besides the actual one may be compatible with the starting suppositional state. Indeed, the actual world may not even be compatible with it.

This characterization of inference is skeletal and leaves many important details to be filled in. For now, I want to remain neutral on as much of the metaphysics of inference as I can for the applications I seek in Chapter 3. But one question will require a little added commentary now: What does reliability come to? That is, what does it take for a range of worlds compatible with the information in an information bearing state to be 'safe' for an inferential transition?

These are especially vexing questions because of how tricky it is to account for the conditions on good ampliative inference. Fortunately I can sidestep the hard part of the question for now: my primary interest is in deductive logic, hence with deductive inference. Accordingly, from here on out, talk of inference will be talk of deductive inference unless otherwise indicated.[12] And the safety conditions on deductive inference are simple: they clearly involve 'maximal safety.' In other words, deductive inference aims at the preservation of correctness (or truth) at *all* worlds compatible with starting acceptance states.

Note that the worlds *not* compatible with an acceptance state are precisely

[12] Is deductive inference a special *kind of inference* which fails in its function even if it is good by ampliative standards, or is deductive-correctness merely a *kind of correctness*, not corresponding to any distinct kind of inference, but rather carving out one of several ways in which inference (*simpliciter*) can count as correctly performed? For simplicity and consistency, I will speak of deductive inference in the former sense, though not much hangs on the issue for now. I explore the relationship between deductive and ampliative inference, taking a more controversial stance, in Chapter 6.

those at which it is incorrect. This means that the condition that good inference preserve correctness at all worlds compatible with starting acceptance states is equivalent to a simpler condition: *preserving correctness at all worlds*.[13]

This concludes my initial sketch of inference and its attendant standard of goodness. To summarize: once we have information-bearing, correctness-governed states that represent how the world is or merely might be, it is helpful to have a kind of mental activity whose proper function is to appreciably and safely extract implicit informational commitments of those states. Inference is the mental activity effecting such an operation. Inference counts as good (*qua* inference) when it successfully fulfills this role—its proper function.

Now, how can deductive logic contribute to an investigation of good deductive inference, so construed? Recall that there are two features of good deductive inferences: (a) they preserve truth at all possible worlds and (b) this fact is appreciable to the inferrer.

As many have noted, condition (b) is not easily subject to systematic investigation owing to its psychological variability. A Ramanujan may inferentially flit through the space of mathematical possibility, correctly seeing 'obvious' steps in ways that baffle the ordinary reasoner. On the other end of the spectrum, as Carroll (1895) notes, it seems possible for the 'phenomenally obtuse' to be unable to see even the most elementary of acceptable inferential transitions. What this means is that appreciability is not obviously subject to systematization without psychologically arbitrary stipulations.

By contrast, the first condition on good deductive inference, (a), ends up being a simple modal property of mental content—that is, a modal property of the truth-evaluable propositional objects of mental states—completely divorced from contingent psychology. To see whether a transition between contents preserves truth at all worlds, we need know nothing about the mental states that bear these contents, or the psychology of the reasoner who is in these mental states. We could hardly hope for a condition more amenable to systematization. And what is more, we can exploit the expression of contents in language to facilitate the process of tracking the relations among those

[13] I've formulated this proper function of deductive inference in terms of correctness, but I could have equally formulated it in terms of (truth-conditional) information: deductive information aims at a total information-preserving transition between acceptance states. An inference should move one to an acceptance state that 'rules out' no more worlds than the states with which one began. This is, again, readily seen to be equivalent to preservation of truth at all worlds.

contents that are of interest to us. That is, if we presume that the propositional contents of mental states are also expressed by the declarative sentences of a (formal or natural) language, we can exploit the compositional structure of such sentences in systematizing necessary truth-preserving relations among such contents.

Now, to investigate relationships of necessary truth-preservation that hinge on compositional regularities would be to investigate *general entailment relations* in compositional semantic theorizing. While this kind of investigation clearly could have an important bearing on the study of inference, the entailment relations investigated in compositional semantic theorizing subsume necessary truth-preserving relations that are not traditionally conceived of as logical. So-called lexical entailments (like the entailment from something's being a vixen to its being a fox) are examples.[14]

Accordingly, to get from here to logical inquiry as more traditionally conceived, we make a broadly familiar move: we note that sometimes sentence transitions express contents that can be seen to necessarily preserve truth 'in virtue' of some restricted set of their linguistic properties.[15] For example, an entailment between sentences (e.g. from "a is red and a is round" to "a is red") may be guaranteed merely under the assumption that its predicate denotations (those of "red" and "round") belong to some general class of predicate denotations (including, e.g., "green" and "square"), rather than bearing the specific denotations they do. If so, a logic may abstract from the details of particular predicate denotations in tracking entailment relations, instead holding constant the denotations of certain special 'logical' vocabulary (in this case "and"). This enables us to use regularities in language to track entire classes of good inference ("a is red and a is round" to "a is red"; "a is green and a is square" to "a is green"; etc.) by a shared linguistic form ("a is F and a is G" to "a is F"). In this work, I will not take a stand on whether the vocabulary held constant in tracking good inference is privileged in any particular way, and so will leave open whether, for example, the semantic properties of "knows", "is true", or even "bachelor" could count as logical.

So, if one starts with a concern for understanding the goodness of deductive inference, this can lead one to investigate the modal properties of propo-

[14] See Glanzberg (2015) for a helpful discussion of such entailments and their relation to logic.

[15] Cf. Sánchez-Miguel (1992), Etchemendy (2008).

sitional mental content. And these modal properties play a supporting role in good inference that can seemingly be divorced from a study of mental states themselves. Finally, since the mental content appears to be expressible in language where the relevant modal properties can be systematically tracked, this in turn leads naturally to an investigation of the expression of content through language. In particular, the goal of systematization can drive us to consider how patterns in necessary truth-preservation among contents are secured by subsets of the linguistic properties of the sentences that express the contents. After all, each such subset of properties defines a class of sentence transitions, and so a class of potentially good inferences, which would share a common linguistic 'form.'

This accordingly provides the conception of logic that I will be exploring in this book:

> *A logic investigates relations of necessary truth-preservation among sentences' assertoric contents in virtue of some subset of their linguistic properties, in order to gain an understanding of a necessary condition on good deductive inference*.

As noted in Chapter 1, it may be methodologically helpful to provisionally see this construal of logic as stipulative. *If* there is a mental activity of inference of the sort answering to my functional specification, *then* the characterization of logic just given provides us with a useful theoretical activity in which we could engage. Whether this activity actually answers to what has gone under the heading of "logic" in core tradition obviously remains to be seen.

So far, this characterization of logic remains a sketch. Let me flag one noteworthy feature of the characterization—the *indirect* character of logical investigation—before reviewing the many important ways in which the characterization is skeletal and *incomplete*.

On the view of logic just sketched, we investigate language in order to investigate facts about assertoric contents, and these in turn help us to understand features of corresponding mental contents that underpin good inference. As such, logical investigation can often proceed in principle without doing much by way mentioning the purposes or implications of its results for inference. Logic as I've described it is concerned with relations among linguistic contents, but only instrumentally, for purpose of shedding light on the goodness of a mental activity. It is in this sense that logical investigations are

inherently, and in some ways unusually, *indirect*. This might lead us to wonder: Once we turn our attention to facts about language or linguistic assertoric contents, what is the point of saying that logic studies inference anymore? Why not say that logic is the study of necessitation relations among contents or of facts about language and leave inference out of it?

An analogy can help to clarify the situation. Consider a hedonist rule utilitarian who thinks that morally correct actions are those which conform to rules whose faithful general adoption would maximize pleasure and minimize pain. When this theorist wishes to turn their attention to the applied question of what actions are *actually* morally correct, they will engage in investigations which do not straightforwardly belong to moral theorizing proper. They may investigate: psychology or physiology to understand how pleasure and pain contingently arise in human or animal subjects; game theory to understand how the adoption of certain rules pans out in settings with multiple interacting agents; and so on. These investigations will help them understand which rules are actually conducive to the maximization of pleasure and minimization of pain. Accordingly it will help them, through their ethical theory, to get information about which actions are morally permissible or not. But note that all of this theorist's ancillary investigations could be undertaken by other theorists for wholly independent reasons having nothing to do with ethics. The physiologist, for example, need not believe hedonism to study the physical basis of pleasure and pain. Foundational ethical questions need never have occurred to them. Likewise for the game theorist.

In a similar way, a metaphysician may study necessitation relations among contents and a linguist may be interested in how, compositionally, we can track them. And philosopher of language may be interested in both of these things. But what distinguishes the logician from these other theorists is the *purposes* for which they investigate these matters, just as with the hedonist rule utilitarian above. In both the ethical case and the logical case as I've described it, we tie target facts of inquiry (inference, ethics) to facts from a different domains that can be investigated independently (metaphysics, linguistics, physiology). In the ethical case, I think this is not liable to mislead anyone. No one seeing the hedonist rule utilitarian at work would think that animal physiology, say, represents their 'ultimate' target of investigation. But some philosophers *have* talked as though logic studies abstract entailment relations, or linguistic

properties, without doing so in service of some further end.[16]

Why would the purposes of one's inquiry matter? For example, when the logician and the linguist (say) investigate the compositional behavior of a language, what difference will it make that the logician has a particular application of the results of this investigation in mind? Generally, it will affect the *methodology* of the investigating theorist. It will make a difference to which questions they make their focus, what idealizations they allow, and whether they downplay or simply ignore certain investigative outcomes. As we will see, the logician may be happy to investigate 'formal' languages with a stipulated compositional structure quite unlike that of a natural language. With a language's semantic structure stipulated in this way, it ceases to be a reasonable target for empirical investigation. This creates quite substantial distance between the logician and the linguist interested in the contingent compositional structure of natural languages. Also, the logician's interest in necessitation relations and language is based on *bridge assumptions*—for example the assumption that the mental contents figuring in inference are the very contents expressed in assertion by declarative sentences of a spoken language. A bridge assumption like this can turn out to be an oversimplification. It may not hold true of certain sub-domains of natural language discourse, in which case the logician may regard the behavior of these domains of discourse as irrelevant to their investigations in ways that the linguist obviously should not.

We will see many examples of the importance of these methodological constraints on logical inquiry much later in our case studies of Part II. For now, I simply want to flag that although my logician can be led to investigate domains like linguistics or metaphysics by a desire to systematize, the logician's heart lies elsewhere, and this can matter in subtle ways to the shape of logical investigation.

The applications of Part II are a long way off, however, and fleshing out the sketch of logical inquiry enough to make those applications requires details to be filled in. In this sense the proposal I've made is incomplete. Let me flag some of the most significant questions that need to be addressed.

(A) What kinds of worlds (e.g., nomologically possible, metaphysically possible, 'logically possible,' etc.) matter to the goodness of deductive inference, and so to logic?

[16] See, for example, how Harman (1984, 1986) sets up logic as the study of abstract entailment relations.

(B) What is the nature of the appreciation requirement on good inference, and does it have any impact on logical inquiry?

(C) What is it to 'abstract away' from non-logical linguistic properties of a word or sentence in the process of inferential systematization?

(D) If a logical transition between sentences secures truth-preservation at all worlds in virtue of some subset of linguistic properties, what is the nature of this 'in virtue of' relation (e.g. is this relation semantic, epistemic, or metaphysical)?

Questions (A) and (B) will be answered together in Chapters 4 and 5. There I will argue that a historical problem about the bounds of cognition reveals that the appreciation in good inference involves a *sui generis* cognitive relation which links good deductive inference directly and exclusively to the space of *metaphysical* modality. Question (C) will be addressed in Chapter 7. There, I'll exhibit the processes of abstraction I have in mind by showing how we can use them to extract first-order consequence as a 'genuine form of validity' for a first-order language. In the process, without taking a full stand on the nature of an 'in virtue of' relation in answer to (D), we will gain an appreciation of how and why it should be a *necessitating, asymmetric explanatory* relation of some kind.

For now, however, I will hold off on exploring the answers to any of the questions in (A)–(D). The reason is that even without answering them, we can use the mere skeletal conceptions of inference and logic sketched in this chapter to do useful work illuminating some aspects of logical inquiry. In particular, the skeletal framework already gives us all the elements needed to probe the important foundational issue of how logic could count as a form of normative inquiry. So let's turn to that issue now.

CHAPTER 3

Inference and the Normativity of Logic

The loosely sketched conceptions of logic and inference just given in Chapter 2 make room for a constrained normative role for logic to play. One of the distinctive features of inference is that it is a goodness-fixing kind. And logic provides us with a way of investigating the conditions on the goodness of inference, though in a somewhat limited and indirect way.

In particular, logic…

(i) …*non-exhaustively* tracks a …

(ii) …*necessary* but *insufficient* condition on …

(iii) …an *evaluative* normative status governing…

(iv) …*acts* or *events*.

That is, logic tracks a necessary condition (truth-preservation at all possibilities) on the evaluative status (goodness) of the act of inference. It does so non-exhaustively (notably ignoring the goodness of certain lexical entailments). And the necessary condition it tracks is insufficient for the evaluative status (since logic sets aside psychologically variable appreciability requirements on inferential goodness). How might logic and its normative scope, so understood, fit into existing debates?

To begin, the view would have immediate implications for a burgeoning program in philosophical logic: the task of finding normative 'bridge principles' for logic. The terminology derives from MacFarlane (ms/2004), but the project owes its life to a kind of skeptical challenge dating back to Harman (1984, 1986).

In §3.1, I give an opinionated and critical survey of the state-of-play in the philosophical program providing logico-normative bridge principles. The criticisms I make of this program, it should be stressed, have their force independently of the conceptions of logic and inference laid out in Chapter 2. The remainder of the chapter in §§3.2–3.5 will explore how the idea that logic studies inference reorients our perspective on the normativity of logic, illuminates the struggles encountered in attempts to provide bridge principles, and gives us new resources to overcome those struggles.

3.1 Bridge Principles and their Discontents

In the course of asking how logic relates to good reasoning broadly construed, Harman suggests that logic's role in reasoning would be cashed out with something like the following principles.[1]

> Logical Implication Principle: The fact that one's view implies P is a reason to accept P.
>
> Logical Inconsistency Principle: Logical inconsistency is to be avoided.

MacFarlane (for different theoretical purposes) generalizes the idea by introducing a schematic form for such principles, dubbing their instances *bridge principles*. These forms link facts about logical entailment and normative claims about attitude states roughly as follows:[2]

> If $A, B \models C$ then [normative claim about believing A, B, and C]

MacFarlane suggests that we can view the space of such bridge principles as generated by varying the "type of deontic operator" (he considers: obligations, permissions, reasons), their "polarity" (whether they prescribe believing, or not disbelieving), the scope of their deontic operators, and whether or not we require knowledge of an antecedent for the consequent to hold. After surveying 36 formulations, MacFarlane tentatively seems to endorse the following.

(wo-) If $A, B \models C$ then you ought to see to it that if you believe A and you believe B, you do not disbelieve C.

[1] Harman (1986, 11).
[2] MacFarlane (ms/2004).

(wr+) If $A, B \models C$ then you have reason to see to it that if you believe A and you believe B, you believe C.

The proposal has spawned a number of competitors (or supplements, depending on the theorist). Field (2009) extends bridge principles to apply to credences, eventually proposing a generalization of the following principle:

(D*) If it's obvious that $A_1, \ldots, A_n \models B$, then one ought to impose the constraint that $P(B)$ is to be at least $P(A_1) + \ldots + P(A_n) - (n-1)$, in any circumstance where $A_1, \ldots, A_n$ and B are in question.

Steinberger (2019a), who focuses on finding bridge principles that could encapsulate logic's guidance of good reasoning, suggests the following.

(S) If according to S's best estimation at the time, S takes it to be the case that $A_1, \ldots, A_n \models C$ and S has reasons to consider or considers C, then S has reasons to (believe C, if S believes all of the A_i).

When Harman formulated his original two principles, he did so precisely to highlight their susceptibility to counterexample. Indeed, he raised a barrage of such worries, since expanded by other philosophers. In brief, these include:[3]

> Backtracking: A recognized entailment can be a reason to abandon one's starting beliefs if one has sufficient evidence against the consequent.
>
> Bootstrapping: Although in most logics $p \models p$, simply believing p does not guarantee that one has reason to believe p.
>
> Clutter: It is an irrational waste of our cognitive resources to needlessly clutter our minds with irrelevant consequences of our current beliefs.
>
> Paradoxes: It appears rational to respond to some unresolved paradoxical or puzzling situations by maintaining logically inconsistent beliefs and managing them responsibly in the interim. Cases include: the preface paradox, in which one can reasonably believe of any large class of one's beliefs that one of them is false on broadly probabilistic grounds;

[3] See Harman (1984, 1986), Broome (1999) MacFarlane (ms/2004), Steinberger (2019a) for more extended discussion.

> the Liar or Sorites paradox, in which a small set of highly cherished principles seem to lead directly to contradiction; and cases of conscious anti-expertise, where one recognizes that some proposition p is true just in case one fails to believe p (or fails to know it, or fails to have high credence in it).
>
> Excessive Demands: We are not irrational for failing to believe logical entailments of our beliefs that are sufficiently hard to recognize. Nor are we irrational for holding logically inconsistent beliefs when the inconsistency between them is sufficiently hard to recognize.

All these problems apply pressure to Harman's Logical Implication Principle. And Excessive Demands and Paradoxes seem to undermine even the Logical Inconsistency Principle.

Harman's reaction was to embrace the defeasibility of his principles. But the other philosophers I've mentioned were instead emboldened to revamp them. The revamped principles may improve on Harman's in some ways. But often enough, it is unclear how they even escape Harman's original worries.

Take MacFarlane's (wo-).

(wo-) If $A, B \models C$ then you ought to see to it that if you believe A and you believe B, you do not disbelieve C.

This is essentially a carefully worded version of Harman's Logical Inconsistency Principle, which ran into trouble with Paradoxes. How does the new principle avoid those worries?

MacFarlane focuses on the Preface Paradox, discussing two strategies for safeguarding (wo-). Let's begin with the first strategy, which is to admit the existence of conflicting rational norms. On this view, in the Preface Paradox, while it may be true that one ought to maintain one's current inconsistent set of beliefs (since they are duly responsive to the evidence), it is *also* true that one ought to revise those beliefs to render them consistent. Of course, one can't fulfill both these obligations. But, as MacFarlane notes, the existence of conflicting norms is something we are familiar with from other domains (notably the legal).

There are two ways of understanding this proposal. On the first, the conflicting norms are both norms of subjective rationality; on the second, at least one norm (probably the logical norm) is a norm of objective rationality.

The option on which the norms are subjective simply appears false. Not (or not merely) because it posits conflicting subjective rational obligations. Rather, it is simply one of the two obligations it appeals to—namely, that one is under a subjective rational obligation to change one's beliefs to render them consistent—does not intuitively obtain. The only motivation for positing the obligation seems to be the *ad hoc* grounds that it would make our logical norm general. Even barring this worry, it is hard to overstate the cost of taking both of MacFarlane's proposed norms to be subjective. All of us are, on brief reflection, in the circumstances of a preface paradox with respect to some large class of our beliefs. The proposal on the table is that virtually every single reflective agent that has ever existed was *inescapably* subjectively irrational. It is not merely that the proposal jettisons intuitive subjective epistemic ought-implies-can principles (which many theorists, myself included, would already view as a high cost). Rather it makes subjective rationality so demanding as to be transparently unobtainable. What is more, we are entertaining this cost in an effort to secure the indefeasible subjective rational force of logic. It is surely a pyrrhic victory when we have secured the inescapability of logic's norms by rendering norms of their type essentially impossible to follow.

The second way of construing MacFarlane's first suggestion, on which logical norms are more objective, fares better in all these respects. On this view, one 'ought' to change one's beliefs in a preface paradox case only in a sense that somehow takes into account features that goes beyond one's current epistemic limitations. But problems arise when we ask in what particular way the norm would be more objective. It cannot be that one ought to change one's beliefs if only one could *reason better*. A characteristic feature of the Preface Paradox is that it persists even in the face of completely idealized capacities for reasoning. This seems to leave only one alternative: that the norm is tracking what one ought to believe if only one had *more evidence*. The claim that one 'ought' to change one's beliefs given sufficient added information is plausible. What is implausible is that this is the form that the norms of logic take. It may be an open question whether logic has normative force. But its having force only for those fortunate few with enough evidence to skirt all Preface Paradox (again, no actual ordinary agent will qualify) seems about as good as having no force at all.

Let's turn to MacFarlane's second, alternative approach to the Preface Paradox. This second strategy begins by insisting, counterintuitively, that one

ought to render one's beliefs consistent in preface paradox cases. But we soften the counterintuitive character of this proposal by noting one way that almost anyone could satisfy such norms: by seeking further evidence. Though one currently has inconsistent beliefs, it is plausible to think that there is evidence that would resolve it. Perhaps logic instructs you to try to find it.

I worry that this suggestion confuses epistemic with practical normativity: I'm not sure practical action like reading a book is ever a way to satisfy epistemic requirements of the form we were seeking with bridge principles. But there is a much simpler concern. The strategy MacFarlane proposes isn't general. There are cases of the Preface Paradox where no further evidence exists and this is known. Suppose you have written a lengthy book on sea turtles, but you are the sole individual rescued by aliens shortly before the Earth's destruction in the crossfire of intergalactic war. Earth, its inhabitants, and the legacy of its sea turtles have been vaporized. Even if you believe you have some mistaken belief about sea turtles, there is no obligation, subjective or objective, to seek out further evidence to rectify the situation. That would be a tremendous waste of time given that you know there is no such evidence remaining.

So far I've been arguing that MacFarlane's principle runs headlong into the very kind of concerns Harman raised for it, in spite of the inventive ways MacFarlane suggests to avoid them. And this after considering only *one* of the many kinds of tricky epistemic scenarios falling under the heading of Paradox. This is important to bear in mind when we turn to consider, say, Field's bridge principle applying to credences.

(D*) If it's obvious that $A_1, \ldots, A_n \models B$, then one ought to impose the constraint that $P(B)$ is to be at least $P(A_1) + \ldots + P(A_n) - (n-1)$, in any circumstance where $A_1, \ldots, A_n$ and B are in question.

Field's proposal has the obvious virtue of handling the preface paradox neatly, and in a familiar way: by having a slightly reduced degree of confidence in each member of a large set of propositions, our logical norms can license a very low credence in their conjunction. But it is not clear Field's principle can cope will other problems raised by strange and puzzling epistemic circumstances.[4]

[4]Though I don't have the space to discuss the matter here, I suspect Field's norm sometimes runs into problems with cases of anti-expertise, especially in light of work by Caie (2013) which extends the challenges they raise to the credal setting. Caie argues that provided devices of self-reference are available to generate cases of anti-expertise, an agent who is somewhat sensitive to her own credal states and aware of her anti-expertise cannot have credences satisfying

But let me set this concern aside for now in favor of a simpler issue. Suppose Field's principle is true and general. What is it illuminating about the normative role of logic in particular? Note, for example, that if we take out the logical relation of entailment, and simply replace it with a conditional, we get a principle which looks just as plausible, and is apparently quite a bit more general.[5]

($\overline{D^*}$) If it's obvious that if $A_1 \wedge \ldots \wedge A_n$, then B, then one ought to impose the constraint that $P(B)$ is to be at least $P(A_1) + \ldots + P(A_n) - (n - 1)$, in any circumstance where $A_1, \ldots, A_n$ and B are in question.

But this principle doesn't seem to have anything special to do with logical consequence. We can see this even more clearly when we take what Field's principle is telling us about the normativity of logical truth, (D^-), and compare it with a principle that replaces talk of logical truth with talk of simple truth, ($\overline{D^-}$).

(D^-) If it's obvious that $\models B$, then one ought to impose the constraint that $P(B)$ is 1.

($\overline{D^-}$) If it's obvious that B is true, then one ought to impose the constraint that $P(B)$ is 1.

Again, the latter principle seems about as plausible, while being significantly more general and more fundamental: if we should assign high credence to obvious logical truths, it seems plausible that this is *because* they are obvious truths.

We should concede that Field's proposal is one possible story about the normativity of logic. On that proposal, logic is normative simply because (actual) truths are, at least when their truth is obvious. If something is obviously true, you should probably be confident in it. And if that obvious truth is a truth-functional compound, that has implications for how you confident you

the probability axioms. Field's logical norm doesn't strictly speaking require one's credences to satisfy the probability axioms. But neither does Caie's impossibility proof strictly speaking require the probability axioms to get up and running. The dialectical situation is a little complex: depending on what logic one endorses, one may get pressure from Caie's proof to admit the defeasibility of (D*).

[5] Note that even though we are forced to group the premises into conjunctive form, our principle still gets a non-trivial, and intuitively correct, verdict on the relation between credences in conjunctions and credences in their conjuncts, provided conditionals with the same antecedent and consequent are obviously true.

should be in its truth-functional constituents. Logic is then normative because, in furnishing us with validities and consequences, it thereby furnishes us with actual truths, including some truths in conditional form.

I don't want to disagree with this story. But I think that if it were *all* to say about the normativity of logic, it would be disappointing. It's customary to treat logical truths as true, and very common to take logical consequences to deliver true conditionals.[6] As long as truths (perhaps the obvious or clear ones) have some role to play in regulating belief, logic will. But this would make logic as distinctively normative as any other domain of inquiry that has some 'obvious' truths in it. And it would mean that the methodology of current theorists is getting things backwards: we should start by figuring out what the normative implications of truth, or recognized truth, are, and derive those for logic as a byproduct.[7]

This concern arises for Field in part because of the rider that logical truths be 'obvious' to avoid certain worries from Excessive Demands. I think a related kind of concern may arise for Steinberger's principle (S).

(S) If according to S's best estimation at the time, S takes it to be the case that $A_1, \ldots, A_n \models C$ and S has reasons to consider or considers C, then S has reasons to (believe C, if S believes all of the A_i).

Again, to the extent that (S) is plausible, it is unclear what is special about the logical character of the norm. Norms like ($\overline{\text{S}}$) seem about a plausible, and more general.

($\overline{\text{S}}$) If according to S's best estimation at the time, S takes it to be the case that $A_1 \wedge \ldots \wedge A_n \rightarrow C$ and S has reasons to consider or considers C, then S has reasons to (believe C, if S believes all of the A_i).

[6] Note that this is not presupposing a deduction theorem, but only that consequence delivers *actual* true conditionals. Still, this too, can be doubted: as Field (2009, 2015) notes, even this condition can be violated given certain theories of the liar. My suspicion is that this is simply a count against such theories. But even if not, it is far from clear in these special cases whether, say, my proposed analog principles ($\overline{D^*}$)/($\overline{D^-}$) are doing any worse than Field's, precisely because of the controversies over the rational way to respond to liar-like phenomena.

[7] This is, I suspect, close to an original objection of Harman, who noted that any force of logic for reasoning comes from their being known or recognized, at which point it becomes hard to see how their import for reasoning differs substantially from any other non-logical principles. See also the objections of Russell (2017) and Blake-Turner & Russell (2018) discussed in n.24.

We may again see this a little more clearly by considering the principle's implications for logical truth given by (S$^-$), and a counterpart norm for truth ($\overline{\text{S}^-}$).

(S$^-$) If according to S's best estimation at the time, S takes it to be the case that $\models C$ and S has reasons to consider or considers C, then S has reasons to believe C.

($\overline{\text{S}^-}$) If according to S's best estimation at the time, S takes it to be the case that C is true and S has reasons to consider or considers C, then S has reasons to believe C.

It might be objected that in this case there is an asymmetry. ($\overline{\text{S}^-}$) plainly licenses a form of bootstrapping and to that extent is implausible: it says that if I think p is true, in my best estimation, then I have reason to believe p. I am inclined to agree that ($\overline{\text{S}^-}$) is implausible on these grounds. It's just that Steinberger's (S$^-$) is equally implausible, licensing a related form of bootstrapping: it says that if in my best estimation p is a logical truth, I have reasons to believe p. But this is intuitively untrue.

Donald believes that there is an even number of stars, but quickly realizes he has no information bearing on the question. So he reflects further: is it a logical truth that there is an even number of stars? Donald, who is horrendous at logic, thinks on this matter and (in his best—i.e. horrendous—estimation) settles the question in the affirmative. Does it follow that Donald does in fact have some reason to believe there is an even number of stars? I think not and, accordingly, that (S$^-$) (and so (S)) are wrong: the fact that something holds in one's best estimation provides no reasons at all if one's best estimation is awful. Once we see this, we can see that actually some kinds of bootstrapping worries afflict all of ($\overline{\text{S}^-}$), (S$^-$), ($\bar{\text{S}}$), and (S) equally.

Steinberger may be happy to embrace this consequence. He bills his principle (S) as only supplying a 'directive' norm, which has "the purpose of providing first-personal guidance in the process of practical or doxastic deliberation."[8] He stresses that as such, these norms should only be held to standards consistent with their serving a fruitful role in guiding reasoning. As he puts it: "It may be that the only norms sufficiently transparent to us [to be followable] are ones whose triggering conditions appeal to an agent's states or attitudes."[9] I agree that some conditions on attitudes may be appropriate for

[8] Steinberger (2019a, 316)
[9] Steinberger (2019a, 317).

directive norms. But I think that even if we restrict our attention to directives, Steinberger has still gone too far. Many instances of consequence and failures of consequence are not only *a priori* but blindingly obvious. Nothing prevents reasoners from holding the correct logical views on such matters beyond their logical obtuseness. To be a norm that is follow*able* in the intuitive sense doesn't mean that one has to actually take it to hold (whether explicitly or tacitly)—intuitively it need only be within reasonable epistemic reach. So as it stands Steinberger's epistemic triggering condition is simultaneously too strong (for allowing Donald to get reasons he doesn't have) and too weak (for failing to condemn equally obtuse reasoners who lack any attitudes towards obvious logical facts that bear on their reasoning).

That is my suspicion. But even if this is wrong, the more important point is the consideration of symmetry: any defense of (S^{-}) from bootstrapping worries is liable to save $(\overline{S^{-}})$—again with the latter being more fundamental and general. If so, there is nothing distinctively normative about logic to be found here.

I wanted to mention one final concern for (S), which is far from a knock down consideration against it, but which will be important for my ensuing discussion. This is that principle (S) is extremely weak and qualified: it does not require conformity with logical principles, but only provides defeasible reasons for doing so (similar concerns apply to MacFarlane's (wr+)). To say that we sometimes have *some* reason to conform to logical norms is an extremely weak claim, for it is consistent with those reasons constantly being defeated. (One could, for example, claim that the goodness of lacrosse is a reason for everyone to play it. Reasons supplied in this manner would be trumped in the grand majority of cases.)

One reason for Steinberger's retreat to reasons-based norms is that those have the virtue of giving resources to respond to the Preface Paradox: logical reasons for inferring a large conjunction of one's beliefs from the individual beliefs may be trumped by inductive reasons for humility. But accommodating some defeasibility highlights the nebulous character of the principle. Suppose I have testimony from two reliable sources about the identity of the culprit of a crime, and I believe each. But their recommendations diverge, and I have seen footage convincing me there is a single culprit. My beliefs are inconsistent—so I know from (S) that I have reason to abandon at least one of them. But: sufficient reason? Intuitively, yes. But nothing in (S) guarantees this.

Steinberger (2019a, 323–4) acknowledges this concern, suggesting that we can explain our intuitions using the idea that we have competing logical and epistemic norms, ranked by priority in ways that shift with context. In the Preface Paradox, broader epistemic norms outweigh logical ones, but the reverse holds in 'ordinary' cases like the one I just supplied.

Even if true, this seems more like a description of a desired solution to a problem for logic's normativity than the solution itself. Even if we get an extensionally correct theory, why are the reasons provided by logic varying in strength or efficacy? Is it that they are constant in strength, while the strength of other epistemic norms varies? Or does the strength of the reasons supplied by logic itself vary? Either way: what is that strength? All (S) tells us is that there is some. Being told that it is always enough, in interaction with other reasons, to account for our intuitions can feel dissatisfying. It is true that it is hard to find objections to such a theory. But there are real concerns that this is only because of how non-committal the theory has become.

This last point may seem like an unfair objection. But even if so, I think it helps illustrate an important general point about our trajectory: the program of supplying bridge principles has been marked by continual process of weakening and hedging. As we proceed from Harman through MacFarlane, Field, and Steinberger, operators are given wide scope, epistemic constraints of increasing strength are built into triggering conditions for norms, norms are weakened from strictly obligating to reason-providing. While it is true that we may be getting closer to a true principle through such hedging, there is a concern that the weak principles we arrive at are diluting or omitting something essential to logical force, which is intuitively absolute and exceptionless. We have strayed very far from that initial guiding idea.

I want to press the concern that the bridge principles so far examined have lost sight of something integral to logic's normativity with a final simple set of objections to all of them.

Consider someone who makes a series of counterfactual suppositions and then 'under supposition' affirms the consequent. They suppose q, and if p then q, then conclude under supposition: "well, therefore p." They do nothing further. This, I take it, exhibits a paradigmatically illogical form of reasoning (affirming the consequent is, after all, standardly described as a *logical fallacy*). If there are logical norms of any kind that govern reasoning, it seems this person should have *already* violated them. They need not arrive at a belief

(for example of some conditional) to have contravened logic's dictates and be pronounced a poor reasoner on specifically logical grounds. It is worth adding that they seem to have done something wrong, by logic's lights, that could have also occurred for belief. If someone affirms the consequent while believing, they have intuitively made a distinctively logical mistake in reasoning, and the *very same* one that was made by the supposer.[10]

This raises an immediate, simple set of concerns. First, no bridge principle we have seen has any implications for supposition states. They only speak of beliefs or credences. Moreover, the principles do not seem to be extensible in any straightforward way to suppositions. Take, for example, MacFarlane's (wo-) which essentially forbids believing logical contradictions. The analog of such a principle for supposition is implausible if we are rationally permitted to suppose contradictory information for the sake of a *reductio*—a procedure which often appears highly rational. We do not need anything as recondite as the cases of Paradox to make this point. I think similar things can be said of the other principles. (For example, is it clear that someone is *irrational* if they don't extend their suppositions logically, even when those extensions are under consideration?)

Second, even if the principles were extensible to provide plausible norms for suppositions, none of them seem like they are in a position to say what has gone wrong with either the supposer, or even the believer, who has fallaciously affirmed the consequent. For example, even the believer who affirms the consequent is not necessarily condemned by any of (wo-), (wr+), (D*), or (S): the only principles from this set that forbid anything merely forbid broadly inconsistent sets of attitudes, which our reasoner need never have.

To be clear on one point, the authors I have been discussing are typically not looking for norms that would govern suppositions, and are sometimes quite explicit about this. MacFarlane, for example, seems to acknowledge a possible task for logic that would have implications for mere supposition but, for reasons I will discuss critically in §3.3, claims that task is worth setting aside for a more theoretically fruitful focus on belief.

However, my example raises concerns for that methodology. When one person affirms the consequent under supposition, and another does it while

[10] Authors such as Hlobil (2015) have used examples like this to argue that there are characteristically diachronic norms of reasoning. My use of the example here is for sightly different purposes, which are compatible with taking norms to be synchronic. See the discussion at n.38.

believing, they intuitively make mistakes of the very same, distinctively logical kind. Logic impugns both believer and supposer as bad reasoners, and on the same grounds. What this suggests is that the guiding methodological assumption should be that logic applies to the domains of the supposed and the believed equally, so that what is fundamental and distinctive of logical normativity is felt equally in both. We should of course concede that logic may have special, particular downstream effects for belief as a result of the interaction between logical norms and norms specifically governing belief (like that one should believe obvious truths, or respond well to one's evidence, and so on). But a concern for all the bridge principles seen so far is precisely that they are *mixing* logical and broader epistemic norms, and as a result confusing what is distinctive of logical normativity. This seems all the more apparent from the fact that many clear instances of logically fallacious reasoning, even for belief, are transparently ignored by all the principles we've considered. Something has gone wrong.

3.2 Inferential Goodness and Bridge Principles

The discussion of §3.1 has been quick and my objections there are hardly decisive. There is not only room to rebut my objections head-on, but to continue to refine bridge principles in response to them. My goal so far has primarily been to raise suspicions—to remind that the ways in which bridge principles encounter obstacles, and are successively weakened, give us reasons to think not merely that we haven't yet found the right one, but that there is something misguided about the shape of the project as currently conceived. In this section, I want to deepen and defend that suspicion.

Note two presuppositions built into the form of every bridge principle we've encountered. First, they all involve non-evaluative terminology (*ought*, *may*, *reason*). Second, they apply this terminology to combinations of mental states (*beliefs*, *credences*). It is noteworthy that alternatives to principles of this form are hardly considered in the space of options.[11]

These presuppositions should seem especially noteworthy in light of the

[11] Steinberger (2019a,b) is a rare exception in pointing out that bridge principles can be evaluated along different normative dimensions: as directives, evaluatives, or appraisals. What is intriguing is that Steinberger takes even what he calls "evaluative" bridge principles to be formulable with terms like *ought* or *reason*. This may be connected with the related assumption that the norms apply to agents, or their attitude states, and not acts.

proposal in Chapter 2. If that proposal is on the right track, logical norms are fundamentally formulated in evaluative terms like *good* or *correct*, and the norms apply not to states, but *acts* or *processes* that mediate between them. We could, if pressed, formulate the view using a kind of bridge principle. It would look something like the following:

(Good) If $A, B \models C$, any inference from A and B to C, in which the inference's necessary truth-preserving character is appreciated by the inferrer, is good *qua* deductive inference.

Let me begin by noting two things about this principle. First, it is exceptionless, simply sidestepping all the major concerns for bridge principles including those I newly raised. Indeed, I claim the principle is more or less trivially exceptionless—it is not obvious what could count as an objection to the principle, at least provided the views of inference outlined in Chapter 2 are correct. Second, the correctness of the principle would illuminate why the counterexamples to rival principles are arising in the form that they are: the counterexamples are all the characteristic result of trying to shoehorn a fundamentally evaluative notion governing acts into norms governing something like the act's performance.

To see the exceptionlessness of the principle (Good), let's quickly run through the standard concerns for bridge principles, where we will see a pattern emerge.

Backtracking tells us an entailment can be reason to abandon one's starting beliefs. (Good) is silent on this issue, and so is compatible with this claim. (Good) tells us one way to infer well. It does *not* tell us when it is a good time to infer. If someone has transparently false beliefs, and applies Modus Ponens to arrive at an even more absurd conclusion, the problem is not that they have made a bad inference (that is: an inference performed badly by the standards inherent to inferring). Rather, they have made a correct inference when the situation didn't call for an inference. They have performed an act, well, that they shouldn't have wasted their time performing. (Compare: someone can bake a cake well at a time when they should not be baking—say, they are on the verge of being consumed by a fire.)

Bootstrapping tells us that although $p \models p$, believing p provides no reasons for believing p. (Good) is silent on this issue, and so is compatible with this claim. All (Good) says that if $p \models p$, an appreciated inference from p to p

will be a good inference (which is true).

Clutter tells us it is irrational to clutter our minds with needless entailments of our beliefs. (Good) is silent on this issue, and so is compatible with this claim. If someone clutters up their mind by adding disjunctions to their beliefs (Good) may say that they are inferring flawlessly. It will not say that it was a good idea for them to waste their time performing those, otherwise flawless, inferences. (Imagine a master baker, obsessively baking flawless cakes at the expense of their health or hygiene while earlier cakes they have made begin to decompose. There is a problem with the baker, but it is not that they are baking cakes poorly.)

Excessive Demands tells us that we can't fault reasoners for failing to derive far-flung consequences of their current beliefs. (Good) is silent on this issue, and so is compatible with this claim. Indeed, the far-flung consequences of our beliefs are precisely those we cannot easily appreciate. The motivating theory behind (Good) tells us we *can't* correctly infer those things directly.

Paradoxes tells us that sometimes, in hard cases like the preface paradox, the liar paradox, the Sorites paradox, or cases of anti-expertise, it is rationally permissible to have beliefs not closed under simple entailment, or even to have inconsistent beliefs. (Good) is silent on this issue, and so is compatible with this claim. The form of reply here is as for Backtracking. To take the preface paradox: (Good) only tells us that inferring a conjunction from its conjuncts (in an appreciable way) is to make a good inference. It does not tell us that it is good (rational, permitted, required) to make that inference.[12]

Finally, (Good) is part of a broader view of logic and inference which handles my own objections to bridge principles from the cases of the believer and the supposer who affirm the consequent. For that broader view also yields conditions under which some inferences will count as badly performed, including the following:

(Bad-Appreciation) If, in a deductive inference, what is appreciated as being a ground of necessary truth-preservation is in fact not, then the inference is bad *qua* deductive inference.[13]

[12] This is of course in no way to provide any *solution* to the preface paradox. It is to say that provided there is some solution consistent with the claim that one can maintain contradictory beliefs in the preface paradox, (Good) will give logic a kind of normative force consistent with that solution.

[13] I frame this principle as if deductive inference is a distinctive type of activity (distinct from

(Bad-Appreciation) handles both the believer and the supposer who affirm the consequent, and handles them both in the same way, as I claimed would be desirable. Both the believer and supposer have performed inferences of the same type: they deductively inferred that p from the claims that *if* p *then* q and q, and on the basis of the claims having that form. Inferences of that form will not generally preserve truth at all possibilities, whether those inferences operate over beliefs or suppositions. Accordingly, that inference type cannot be appreciated as a necessarily truth-preserving inference on the grounds the reasoners have employed. As a result, the inferences of both believer and supposer are bad inferences, and bad for exactly the same reason.[14]

I said (Good) 'sidesteps' the traditional obstacles for bridge principles, and I meant it. As should now be clear, (Good) simply makes no commitments about the issues that are pressing for rival views. I also said that it is not clear what could possibly count as a counterexample to the principle, once the view of Chapter 2 is in place. The same goes for (Bad-Appreciation). The reason is simple: if inference really has as its proper function to appreciably extract information from an information state, and uses of "good" or "bad" are tracking whether or not inference succeeds in that function, then nothing—no paradox, no odd epistemic circumstance, no quirk of reasoning—could possibly stand in the way of (Good) or (Bad-Appreciation) being true. Such things can only influence *when* one should exercise one's capacity to infer, not *how* to properly employ that capacity when it is recruited.

I suspect some will see these as vices rather than virtues—indications that my principles are too non-committal, or simply ignoring the issue we wanted to investigate. I will come back to this concern in §3.3. But I want to set this concern aside briefly to highlight an important lesson. Not only does a principle like (Good) avoid the standard objections to bridge principles, but it shows that those objections are, from a certain perspective, unified. Aside from Bootstrapping, the initially disparate objection types above can actu-

ampliative inference). If there is a single activity of inference, which can be assessed for deductive and ampliative goodness, then the principle should be reformulated in the obvious way.

[14] There are of course other ways for a deductive inference to be bad besides those classified by (Bad-Appreciation). An inference in which premises fail to necessitate the conclusion, for example, will *ipso facto* be a bad deductive inference. In the case of affirming the consequent this *need* not be the case, however (consider, e.g., if p is a lexical entailment of q). So (Bad-Appreciation) gives the most general explanation of the failure in this instance, and also helps to reveal why the mistake is intuitively 'logical' in character—as the grounds (wrongly) appreciated are formal in character. Thanks to Chris Blake-Turner for urging me to get clearer on this.

ally be seen to be of one and the same *kind*. This is why the response on behalf of (Good) is essentially the same for each of those objections.

Why is it that standard objections to bridge-principles have this common form? I want to suggest that this is because the program of finding bridge principles has mistakenly been trying to shoehorn an evaluative notion governing a mental act into a constraint on when one ought, may, or has reason to form certain attitudes. We can see why this project would repeatedly encounter a single and recurring style of objection with the help of an analogy.

Consider any other evaluative notion governing an act or activity, and ask what would happen if we tried to capture the evaluation using bridge principles involving deontic language that applies to the agent's performing the activity. For example, take a fastball pitch in baseball. This is a standard form of pitch which is geared at producing a strike by testing the batter's reflexes. Given this purpose, a *good* fastball pitch is one that has (among other features) high speed and little lateral movement.

If we accept that this is what it is for a fastball to be a *good* instance of its kind, what does this tell us about what pitchers *ought to do*? It is not obvious. Consider trying to capture the evaluative notion with claims like the following.

(ia) If A is [known to be] a good fastball, then one ought to pitch it.

(ib) If A is [known to be] a good fastball, then one has reason to pitch it.

(ia) has obvious counterexamples, even restricting our attention to pitchers on the mound in a game. Perhaps the batter is fantastic at hitting fastballs, but terrible at hitting other pitches. Or, perhaps you can get a strike with your perfectly pitched fastball, but you stand to do even better—getting an out—by throwing to first where a runner is leaning too far off base.

I would be tempted to say that in some of these cases (like the first) you don't have any particular reason to pitch a fastball, falsifying (ib). But I don't need such a strong claim for my purposes. It suffices to note that any reasons you do have to pitch a fastball are weak, defeasible, and do not capture the strength of the evaluative notion of a good fastball pitch with which we began.

We might weaken the principles further by moving from prescription to proscription.

(iia) If A is [known to be] a bad fastball, then one ought not pitch it.

(iib) If A is [known to be] a bad fastball, then one has reason not to pitch it.

Again (iia) has simple counterexamples. Maybe you are a terrible pitcher, but your (admittedly bad) fastball is the best of your bad pitches. Or maybe you're a fine pitcher of fastballs, but you could get a strike now with a poor pitch, and get the added benefit of setting up advantageously deceptive expectations for the next, better batter by doing so. In both cases, your bad fastball is the pitch you ought to make (in the second case, precisely because it is bad).

I would be tempted to say that you lack reasons to do otherwise, falsifying (iib) as well. But again, it will suffice to note that any reasons you do have to avoid pitching your fastball are weak, defeasible, and do not capture the strength of the negative evaluative notion of a bad pitch with which we began.

The pattern here should look familiar. We're seeing that when an evaluative standard governs an act, it is not easy to cash this out in terms of reasons or obligations one has to perform (or refrain from performing) the act. What's more, it is easy to see precisely why this would be the case, and what kinds of counterexamples would arise for any attempt to effect that transition.

The evaluable acts we have considered, whether they be inferences, cake-bakings, pitches or anything else, are called "good" in connection with their associated end or purpose. It is from that end or purpose that the act derives its standard of goodness: the features of the act that promote or secure that end. From this two things follow.

First, the standard of goodness being applied is tracking features relative to that fixed end. As a result, calling the act "good", in this sense, has no implications for whether the act should be performed or not. Saying that a pitch is a good one is not to say you should pitch it. That depends on whether the situation calls for that kind of pitch. Likewise, saying that an inference is good (*qua* inference), or performed well, or correctly, is not to say that one should perform it. That depends on whether the circumstance calls for an act of information extraction. There is simply no tension between saying that such-and-such is what it takes to perform an act-type well, but that the circumstances don't call for that act-type at present.

Second, there is a simple recipe for finding counterexamples to the claim that when an act is good as its act-type one should perform it: find reasons against promoting or accomplishing the act's goodness fixing end. Do you want to find a case where, even though a pitch of a fastball is good *qua* fastball pitch, you shouldn't pitch it? Easy: find a case where you have no reason to try to get a strike merely by testing a batter's reflexes. This can be because the

batter's reflexes are too good, or you have better things to do (like getting an out by throwing to first). Or, more simply, it could be because you reasonably can't perform a reflex-testing pitch well at all.

Do you want to find a case where, even though an inference is good, *qua* inference, you shouldn't execute it? Easy: find a case where you have no reason to perform a total act of information extraction on an acceptance state. That could be because one is in a position to see that basing acceptance state contains false information, and the concluding acceptance state is somehow regulated by truth (Backtracking); or because even though the information in the acceptance state seems reliable, you have more useful things to do (Clutter); or because you're in a tough epistemic situation where extracting the information from your belief state is going to lead to foreseeably incorrect, and otherwise pernicious, acceptance states (Paradox). Or maybe more simply, just find a case where you are (reasonably) not in a position to make the inference well due to your own limitations (Excessive Demands).

So here is the final lesson: if the normativity of logic is as described in Chapter 2, we can see that the standard barrage of counterexamples to existing bridge principles are almost all of a single, predictable form. They are all the very sorts of objections one would encounter if one were mistakenly try to take an evaluative normative notion governing an act, and transpose it to illegitimately draw conclusions about when one ought, or ought not, perform the act.

One could, of course, adjust the resulting norms (whether they concern inference, or pitching, or cake-baking, or any other act) to avoid the counterexamples. But one could only do this at the expense of losing sight of the original norm governing the goodness of the act. The way to do this would be to start encrusting distinct conditions that track the norms governing not, or not only, the goodness of the act, but the conditions which make it reasonable to perform the act. In the case of inference, at least if we fixate (illegitimately) on cases involving only beliefs, these encrusted conditions will start to tack on features that make beliefs reasonable in light of their truth, or apparent truth. And, as I've already noted, this is precisely what we seem to see in the progression of principles building on Harman's starting point.[15]

[15] Indeed, in trying to formulate principles explaining merely how truth governs belief, one finds precisely the same kinds of moves as in the normativity of logic literature: wide-scoping, strengthened epistemic triggering conditions, and various contortions to avoid paradox. One can in fact see all of these moves made in the survey of truth-governed norms for belief in

As such, the importance of the view of Chapter 2 is not merely that it arrives at a conception of the normativity of logic which avoids the existing barrage of counterexamples to bridge principles. Nor is it even that the view seems to stand immune from any similar form of counterexample. Rather, the view illuminates key methodological presuppositions—presuppositions that can and should be questioned—that seem to lie at the heart of the discontents of existing treatments of logic's normativity.

3.3 Logic and Reasoning

In §3.2, I set aside two related concerns that it is now time to take up. First there is a concern that my view of logic's normativity is weak, precisely because it avoids the threat of standard counterexamples by being non-committal. Second, there is the related concern that my view is not properly engaging with any of the concerns about logic that lead philosophers like Harman, MacFarlane, Field, or Steinberger to investigate logic's normativity.

This second concern is especially pressing, since it is unclear whether the authors I engage with even have a common theoretical goal in providing bridge principles.[16] Fortunately, despite the differences between their approaches, there is at least one unifying thread among most discussions of bridge principles: a concern with logic's relevance to *reasoning*. Accordingly, the first step in assessing how the view of Chapter 2 is engaging in the dialectic is to say what implications that view has for reasoning.

We should begin with Harman. Harman, I noted, supplied bridge principles mainly to emphasize their susceptibility to counterexample. As Harman (1986, 5) puts it: a "logical principle [i.e., a logical relation of consequence or what Harman calls an "implication"] holds without exception, whereas there would be exceptions to the corresponding principle of belief revision"—where principles of belief revision are enshrined in bridge principles.[17] For Harman, this contributes to a central contention that logic is "not of special relevance" to a theory of reasoning, where "reasoning" is interpreted broadly to involve

Bykvist & Hattiangadi (2007).

[16] See in this regard the careful and illuminating discussion in Steinberger (2019b).

[17] Harman sometimes calls these latter principles "rules of inference" (Harman, 1984, 108). It should be borne in mind that his "inference" is tracking reasoning broadly construed, not the process I have called "inference" in Chapter 2.

general procedures for revising one's beliefs, including abandoning them.[18]

Why did Harman take the defeasibility of bridge principles to contribute to the claim that logic is not of special relevance to reasoning? Harman certainly does not think that logical rules of implication have *no* importance for reasoning at all. Indeed, he takes them to stand as integral truths that help regulate what we should and should not believe, albeit defeasibly. But as such it is not clear what makes logic more relevant to reasoning than other suitably general and stable truths. For example, as Harman notes in regard to the issue of inconsistency, "[p]rima facie, one should not continue to believe things one knows cannot all be true, whether this impossibility is logical, physical, chemical, mathematical, or geological."[19] This is not Harman's only case against logic's special relevance to reasoning, but it is central.

In §3.2, I may seem to have been siding with Harman on the issue of reasoning, perhaps emphasizing a different role for logic to play. After all, a central idea of my discussion is that logic helps track conditions on when an inference is performed well or correctly, but that doing this should be sharply separated from any claims about *when* to perform an inference—even a good one. It is clear that Harman is keenly interested in the latter kind of question.

The claim that I side with Harman is partially right, but it is worth noting that the dialectical situation is somewhat complex. Let me begin with some ways in which I agree with Harman. First, I've conceded that logic doesn't give information about how to reason, in the sense of when to engage in reasoning of certain kinds. Instead logic has implications for how to correctly perform an act that is part of reasoning—namely, deductively inferring. What's more, in investigating conditions on good deductive inference, deductive logic only investigates one of many such activities of reasoning (including inductive inference), it tracks the goodness of this activity imperfectly (by ignoring lexical entailments), and it only investigates a necessary but insufficient condition on that goodness (by setting aside appreciation). Harman himself emphasized points similar to all these three in attacking logic's special relevance to reasoning.

But when Harman expands on his claim that logic lacks special relevance for reasoning, he sometimes goes too far. Inference, as just noted, is *part* of reasoning. It is a central, if not essential, such part. A reasoner simply could

[18] Harman (1986, 11).

[19] Harman (1984, 109). See also the discussion at Harman (1986, 17).

not get by, as a reasoner, in any ordinary course of existence without the ability to draw deductive inferences. (Imagine an agent seeking food who knows that if the prey didn't go down path A it went down path B, and that it didn't go down path A. But the agent is unavoidably stuck in attitudinal limbo, not merely because they can't see the goodness of the inference, but because inferring isn't even in their mental repertoire.) I think it is even an open question whether, if an agent lacked the ability to draw deductive inferences, we could even consider them a reasoner, or a thinker more broadly.[20]

It is thus highly misleading to portray logic's relevance to reasoning as being like that of, say, physics, chemistry, or geology. It is highly misleading to compare the implications of logic to those like "*X plays defensive tackle for the Philadelphia Eagles* implies *X weighs more than 150 pounds*."[21] Physics, chemistry, and geology, for instance, provide truths that one *reasons with*, as arguably does the inductively supported general truth about Eagles defensive tackles just given. Such truths do not, properly speaking, constrain reasoning processes, but merely furnish the materials for reasoning. The standards of goodness governing inference normatively constrain a process of reasoning itself. An agent with no knowledge of the physical, chemical, or geological sciences, or of American football, can still be an excellent, indeed perfect reasoner. But being unable to draw deductive inferences well could be devastating to any reasoner—possibly even precluding them from counting as a reasoner at all.[22]

[20] These remarks about the *importance* of deduction are controversial—see, e.g., the quotations in n.27. They will receive further defense in Chapter 6 when we consider how deductive and ampliative inference are related. For now, it is worth noting that even if deductive inference is rare, and to that extent less significant, it will not change the force of the *conceptual* charge I am making against Harman here and below—that he conflates the objects and processes of reasoning.

[21] Harman (1986, 17).

[22] Part of what is holding Harman back, at least in his early writing, is that he finds himself unable to rule out a view on which logic merely consists of a body of truths, distinguished at most by their generality. (Harman is clear, at least in Harman (1984) that he also finds himself unable to completely agree with such a view.) He considers against this idea only an argument from Carrollian regress. I am not even sure we need arguments against this view in the current dialectic—I think it is perfectly reasonable to take the rival view that logic tracks necessary-truth-preserving entailment relations as a starting point barring further argument. But if we need argument, we can try to do so from the ground up as in Chapter 2. Alternatively, we can simply attack the rival construal of logic on its own terms—see in this regard especially Etchemendy (1990) who (rightly or wrongly) attributes a view like that Harman discusses to Tarski. Another obstacle is that Harman only engages with the 'acceptance' of logical rules either in terms of belief, or in terms of brute dispositions. But there are noteworthy alternatives (see especially the above literature cited in nn.8,9).

To clarify this point, it may be helpful to develop another analogy. Suppose someone said: "Being able to bluff well is of no *special* relevance to playing poker well." I suspect poker aficionados would take exception to such a claim. But we can concede that there is one understanding of this claim which is true: it's not as if bluffing is all one does in poker. To play poker, and especially to play it well, one often has to non-deceptively play the strength of one's current hand or simply fold.

But suppose our character continued: "What's more, being able to bluff well is of no more relevance to playing poker well than having a good hand." At this point, our speaker has gone beyond a potentially misleading statement into confusion. A poker hand is what one plays poker *with*. One can play well with a bad hand, or play poorly with a good one. To lump the possession of a good poker hand in with the skilled actions of strategically betting or folding shows a serious confusion about the nature of the target of investigation.

In claiming that logic is no more relevant to good reasoning than the sciences, Harman has not merely understated the importance of logic to reasoning, but confused the distinctive way in which logic contributes to that study. Truths are what one reasons with, including through deductive inference, just as a hand is what one plays poker with, including by bluffing. One can reason poorly with truths, and well with untruths, just as one can play poorly with a good poker hand, and well with a bad one. Harman's claims about logic's lack of special relevance to reasoning are, I think, sometimes founded on a conflation of the activities of reasoning with their objects. If we think in this way, we are apt to miss the one way in which logic should be viewed as special for the study of reasoning. As Etchemendy aptly put it in a different but related context: "Logic is not the study of a body of trivial truths; it is the study of the relation that makes deductive reasoning possible."[23]

So although I agree with Harman on many points, I think he sometimes goes too far in trying to downplay the importance of logic in the study of reasoning. Logic has what I would think of as a very significant role in that investigation: it studies a huge class of content-transitions that undergird a necessary condition on performing a central activity of reasoning—deductively inferring—well. It studies, (indirectly, and with certain limitations) how to correctly perform an action that is partially constitutive of reasoning well. Whether this is a 'special' role is a vague matter. But one cannot, as Harman

[23] Etchemendy (1990, 11).

occasionally does, compare this role to knowledge of certain truths, even important and general ones, which is no part of good reasoning to begin with. The problem with Harman's discussion of logic's role in reasoning was not, as many seem to suppose, that he failed to adequately refine his considered bridge principles, but rather that he failed to properly locate the distinctive *kind* of contribution that logic makes to the study of reasoning—a contribution which isn't formulable in terms of principles constraining combinations of attitudes.[24]

What of other authors? MacFarlane's discussion is perhaps the most instructive to consider. For MacFarlane seems to recognize something very close to the normative role for logic that I have set out. But as soon as he notes it, he sets it aside as straightforward and unilluminating.

MacFarlane cites an interest in getting clear on the normative role of logic in reasoning as opening up a way to arbitrate logical disputes, both over choice of logic and over foundational questions in logic. Like Harman, he is careful

[24] I see essentially this conflation also lying behind the contentions of Russell (2017, §4) and Blake-Turner & Russell (2018, §3), who claim that logic is not normative in any interesting sense because core logical statements are descriptive, and only have normative consequences alongside other normative assumptions.

> In conjunction with common normative commitments concerning truth and falsity (only believe what is true, don't reason to false conclusions, etc.), logics . . . have normative consequences . . . if this is how logic is entangled with the normative, then it shares this status with paradigmatically descriptive scientific theories, including those of physics and mathematics.
>
> (Russell, 2017, 10-11)

I agree that if the normative import of logic only comes via norms governing truths, there would be nothing that would set it apart from physics or mathematics. (Indeed, I've argued against several theorists like Field and Steinberger above making essentially that point.) But the problem Russell and Blake-Turner rightly identify is not merely arising (as I think they intimate) from the fact that normative consequences for logic only result from pairing its descriptive claims with some separate normative commitments or other. It is rather the presumption that *normative commitments about truth in particular must be doing the work*. That is what would lump logic in with mathematics and physics. There are other normative commitments, besides those involving truth, that bring out constrained subsets of descriptive truths as having *distinctive* normative import. Suppose, for example, that act utilitarianism is true. Then the fact that such-and-such act is utility-maximizing may be a purely descriptive one, that only has normative consequences in conjunction with act-utilitarian principles. But it would be extremely misleading to say in this context that facts about utility-maximization were therefore normative in a way no different than those of mathematics or physics. On the view I'm putting forward, the primary point of investigating descriptive facts about content through logical theory is that these have distinctive import for the goodness of a mental activity. This import is not shared by the truths of physics and mathematics, and it would accordingly be confused to treat logic's normative status as on a par with theirs.

to distinguish different things one could mean by "inference" or "reasoning". And in making one such distinction he says the following:

> In a more formal sense, reasoning is a process of drawing out the consequences of a given set of premises. One need not believe the premises: one might just be investigating them, or using them in a conditional proof or *reductio ad absurdum*. To distinguish this process from reasoning in the sense of "reasoned change in view," we might call it "inferring" (though "inferring" may be subject to the same kind of ambiguity as "reasoning").
>
> …I think it is relatively uncontroversial that logic provides norms for inferring (in the narrow sense of drawing out consequences). For the proof rules of a logic are *explicitly* normative: for example, the $\supset$-elimination rule says that if you have already written down A and $A \supset B$, you may write down B. These proof rules license or permit certain inferences.
>
> …So here is a clear sense in which logic is normative for reasoning. But this sense isn't going to help us much with the problems we looked at in the last section. Our intuitions about when it is permissible to infer a conclusion from some premises (in the narrow sense) have the same sources as our intuitions about logical validity: primarily, our logical training. (Indeed, it takes some logical training in order to engage in the practice of "inferring" at all: one must be trained not to use information not contained in the premises, for instance, and not to worry about whether the premises are true.) Thus these intuitions are likely to be subject to just the same "indoctrination biases" as our intuitions about validity. A classicist will take it to be correct to infer anything from a contradiction in formal argumentation, while a relevantist will not. If we are to get beyond this kind of conflict of intuitions, we need to talk about norms for reasoning in the broader sense: norms for belief and belief change.
>
> MacFarlane (ms/2004).

To evaluate MacFarlane's claims here, I need to draw an added distinction: that between formal reasoning through the use of a particular deductive system of the sort that is taught in a logic class (perhaps on the added assumption that

it is a 'correct' one) and the mental activity I discussed in Chapter 2. I reserve the term "inference" for the latter, and will call the former "symbolic reasoning".[25] I think MacFarlane could have meant either of these two things by his "inference". Accordingly, it is worth considering how his remarks fare on each interpretation (without saddling him with either one).

Much of what MacFarlane says in this passage holds true of symbolic reasoning. For example, the claim that proof rules are normative in the sense that they preclude certain ways a deduction may proceed, but allow others. Also it seems true that it takes training to engage in the process, and that this training may open up theorists to biases toward certain proof theoretic frameworks. I take issue with none of these claims.

But if we were to interpret MacFarlane as talking about what I call "inference", some of the corresponding claims would be true, while others would be problematic. MacFarlane describes the target phenomenon as a "process of drawing out the consequences of a given set of premises" which is very close to my characterization of inference as an information-extracting act. And as I've emphasized, information-extraction does not necessarily operate on belief states, so that it is not necessary that one believe anything while inferring—again according with MacFarlane's characterization.

But if we read MacFarlane as discussing inference in my sense, it is hardly "relatively uncontroversial" that logic provides norms for it in the form he suggests: as permissions to draw conclusions given premises. In fact, I've argued that this is straightforwardly false: if A entails B, one is no more permitted (rationally, or in any other non-trivial sense) to infer B *ipso facto*, any more than it follows from the fact that a player's fastball pitch is good one that they are permitted (rationally, instrumentally, or even by the rules of baseball) to pitch it.[26]

Also, importantly, the claims of the last paragraph quoted above should be

[25] Cf. a related distinction in Rumfitt (2015, §2.2) between "inference" and "deduction". Though the latter doesn't correspond directly to manipulations of a formalism, it is clear that it is a kind of 'meta'-reasoning—reasoning *about* contents rather than *with* them (since one can 'deduce' a conclusion for Rumfitt without accepting it). Notably, Rumfitt draws this distinction precisely to emphasize that logic *does not* concern itself with inference in roughly my sense and *instead* concerns itself with the distinct activity of deduction. This apparently slight distinction actually leads to significantly different conceptions of logic. To take one point of contrast, Rumfitt is not led to spend much time thinking about the nature of mentality and mental contents, which we will see is central to my investigation.

[26] Note, this is true *even* of inferences under suppositions, due to concerns of wasting time.

controversial if they concern inference in my sense. It is far from clear that it 'takes some logical training in order to engage in the practice of "inferring" at all.'[27] This claim is not clearly true even if we restrict our attention to deductive inference. Such inferences have been made continuously by mathematicians throughout history, not all of whom had particular training in logic. It is also possible such inferences are made by ordinary persons regularly, in simple applications of rules like Modus Ponens. Indeed, I think it is an open question whether animals engage in inferences, and even deductive ones.

What is true is that one needs instruction to *theorize about* good inference. In this respect, processes of inference might loosely resemble, say, the kinds of processing that go on in composing and parsing syntactic structure. This composing and parsing is something we do all the time. And, we can suppose, we often do it correctly. But it may still take hard theoretical work to say precisely what it is to do it correctly.

Are our intuitions about what makes a deductive inference good subject to indoctrination biases? Probably. And in this respect perhaps one could maintain there is much more bias in the logical or broader inferential case than in the case of syntax. But one thing that MacFarlane is stressing is that we somehow improve our situation by discussing broader norms for reasoning. This suggestion, if made in respect to the study of logic's normativity within the sphere of what *I* have been calling "inference", is as far as I can see not only unhelpful, but actually counterproductive.

The suggestion is unhelpful because, as stressed in my discussion of Harman, inference is a constitutive and central component of reasoning more broadly construed. In reasoning, and in order to reason, one must sometimes infer.[28] What this means is that even if we could pinpoint the norms for reasoning more broadly construed, one of two things would be true. Either those norms wouldn't happen reflect the influence of the norms properly governing inference, or they would. If the former, then the norms actually end up telling us nothing pertinent to logic—they concern those aspects of reasoning

[27] Cf. the claim of (Rumfitt, 2015, §2.2) that what he calls "deduction" (notably contrasted with a mental activity Rumfitt calls "inference") is an "intellectual activity …rare in everyday life." See also (Dutilh Novaes, 2020, 200): "deductive argumentation/reasoning belongs to niches of specialists and to specific contexts."

[28] This is even true if we posit, as I do not think is true, that inference only operates suppositionally. For the norms governing the inference under supposition will influence how that suppositional reasoning eventually has downstream consequences for belief.

which have no bearing on logical domains. If the latter, then those norms are bound to be just as controversial, if not more controversial, than the simple normative claims about inference itself. There is no reason to think that when the goodness of a particular belief-forming strategy critically turns on whether a particular inference is a good one, that the belief-forming strategy is going to be any less controversial, or influenced by indoctrination biases, than the inference considered in isolation.[29]

Indeed, this thought not only reveals that MacFarlane's proposal to consider norms of reasoned change in attitude is unhelpful, but in fact counterproductive. Once we move to investigations of reasoning more broadly construed, we will be investigating a mixture of logical and non-logical norms. If we focus especially on norms governing reasoning with full beliefs, for example, those norms will certainly integrate broader norms for belief formation, that may have nothing to do with inference in particular, and so nothing to do with logic. This is precisely what I argued has happened to investigations of the normativity of logic in §3.1: traditional bridge principles, to the extent they avoid counterexamples, start to encrust norms that concern things like proper responsiveness to truths. Precisely how those norms of truth-responsiveness should be formulated, especially in puzzle cases like those in Paradox, is itself a tricky matter. So there is interesting work to be done here. But the more of it we do, the more we obscure the distinctive role of logic.

As I say, it is an open question whether MacFarlane was merely concerned with what I have called "symbolic reasoning", or whether he had an interest in the phenomenon I have been calling "inference" (or yet something else). Either way, his motivation for looking at reasoning broadly construed is flawed. Reasoning broadly construed comprises the phenomenon that logic studies as an integral, but proper, part. Trying to find general norms for reasoned belief revision not only fails to avoid controversies proper to the logical domain, but only obscures the normativity of logic by mixing its controversies with those

[29] One of MacFarlane's key applications is to use bridge principles to arbitrate particular logical rules. Others like Steinberger (2016), have followed suit. For example, Steinberger uses his investigation of bridge principles to diffuse relevantist attacks on Ex Falso. Some of these applications can persist, but on different grounds. For example, we can continue to criticize relevantist attacks on Ex Falso as follows: the problem is that relevantists themselves are illegitimately, sometimes implicitly, invoking bridge principles in their arguments that are not only subject to counterexample, but do not clearly bear on logic proper. The right way to object to relevantists is not to look for better bridge principles, but to note that their use of them to motivate logical restrictions is illegitimate.

concerning belief formation that need have nothing to do with logic in particular. Accordingly, it is unclear what insight into logic could be gained through the suggested strategy.

Similar worries can be raised for the investigations of Steinberger and Field, sometimes in ways that are exacerbated by the concerns of §3.1. Steinberger bills himself as following Harman in exploring the normative import of logic for reasoning in general, with a focus on how such norms could figure as *directives* which, recall, "have the purpose of providing first-personal guidance in the process of practical or doxastic deliberation."

I argued that the norms Steinberger ends up with are frustratingly weak, positing only the existence of some logical reasons, in some very special circumstances, of largely unspecified strength. But the important thing to note is that once we recognize that inference is a component process of reasoning, it is with respect to inference, and not belief, that we should first look for guiding principles that distinctively owe their force to logic. I do not have the space to explore such principles in any detail here, but it may be worth noting the starting point. Some obvious candidates for directives are: "Never infer badly" or "If one ought to infer, make a good inference." Logic, of course, only provides an indirect and limited investigation of the goodness or badness at issue. Still, such principles do not succumb to the problems from Backtracking, Bootstrapping, Clutter, Excessive Demands, or Paradoxes for the same reasons as (Good) and (Bad-Appreciation). And they also likewise extend the directive normativity of logic to the merely supposed.

Finally, consider Field. Field (2015) bills his credal norm as contributing to a "conceptual role" for the term "valid", thereby giving us insight into a kind of common core of the concept that is what is up for dispute in logical debate. As he puts it: "*disagreement about validity (insofar as it isn't merely verbal) is a disagreement about what constraints to impose on one's belief system*."[30]

For reasons noted in §3.1, this is dubious. The constraint given by Field doesn't seem to have any special connection with logic. Recall what his principle tells us about logical truth, and how this compares to plausible constraints on simple truth:

(D^-) If it's obvious that $\models B$, then one ought to impose the constraint that $P(B)$ is 1.

[30] Field (2015, 42), emphasis in original.

($\overline{D^-}$) If it's obvious that B is true, then one ought to impose the constraint that $P(B)$ is 1.

Certainly two people can agree about what the obvious truths are, but disagree about which of them are logical. Surely they will agree that they should have high credence in all the obvious truths, regardless of whether they are logical. These two characters are in distinctively logical disagreement. But I find it hard to see what Field's conceptual role can do to illuminate it.

There is more to say, but let me leave off discussion in order to return to the opening question of this section: is my proposal engaging with those of my targets? What I've tried to argue is that the situation is complex. In one important sense, I am not engaging directly with the concerns of the authors I have discussed. Those authors are all concerned with general norms for reasoned change in attitude states. And I am, by my own admission, providing no such general norms. This is effectively how the view I give is able to effortlessly sidestep the slew of traditional counterexamples to bridge principles.

But in another more important sense, I take myself to be engaging with my targets quite directly. These philosophers are looking for norms governing reasoning on various motivating grounds: to gain insight into the special importance of logic for good reasoning, including its directive normative force; to arbitrate logical disputes; and to find a core concept underlying non-verbal logical dispute. In each case, I've argued, the focus on finding norms for reasoning broadly construed is a mistake. And the source of the mistake owes, I think, at least partly to overlooking the plausible role logic has to play in studying inference and the distinctive role inference itself plays in reasoning.

3.4 States v. Acts: More Problems

I've been stressing the importance of taking logical normativity as an *evaluative* form of normativity governing a *mental act*. In §§3.1–3.3, I stressed that it was the characteristic interaction between these two aspects of logical normativity which gives rise to the distinctive challenges encountered in the normativity of logic literature.

In this section, I want to focus on two added concerns that arise more specifically from a focus on mental states rather than mental acts. This may be important, because many authors note and discuss the possibility of for-

mulating logic's normative importance in evaluative terms.[31]

An illuminating case where we can see the importance of focusing on acts or events rather than states arises in Dogramaci (2015b). Dogramaci helpfully notes that logic texts frequently introduce their subject matter in evaluative terms.[32] He further notes that this evaluative talk is then supplanted with talk of logical validity, typically construed as truth-preservation across a range of cases. So, Dogramaci reasonably asks: What justifies this slide? Why is logical validity, construed as truth-preservation across a range of cases, *good*?

The hard thing to account for, Dogramaci claims, is why non-actual truth-preservation matters. The value of actual truth-preservation is easy to explain. After all, true beliefs seem like they are good or valuable in some sense.[33] And actual truth-preserving inference helps us get from true beliefs to true beliefs, thereby expanding our store of valuable mental states. Indeed, actual truth-preserving inference can be valuable even if we infer from a false belief, since this may lead us to a more obviously false belief, which then leads us to jettison our starting belief of disvalue.

Dogramaci isn't alone in highlighting this kind of role for logical inquiry. In a continuation of the passage cited in §3.3, MacFarlane says: "We engage in [inference] (and train our students to engage in it) not for its own sake, but because we think it is useful for telling us what we ought to believe. We infer *correctly* when we infer in a way that is conducive to this goal."

From such a perspective, it would seem that a key interest in performing inferences well is to further the goal of having true beliefs, or beliefs we otherwise ought to have.

But, Dogramaci rightly notes, valid inferences typically preserve truth across an enormous range of cases, many of which any reasoner would immediately recognize as wildly implausible—not even remote contenders for how things might actually be. Because any remotely competent reasoners would rule these cases out, considering them seems irrelevant if one is seeking to gain new true or rational beliefs. For example, depending on how 'cases' are con-

[31] Notably: Dogramaci (2015b), Steinberger (2019a,b). In other domains, like the search for rational constraints on credences, one finds arguments for privileging the evaluative perspective as well. See, e.g., Titelbaum (2014).

[32] E.g., Barwise & Etchemendy (1999, 1) advert to the term "correct", Salmon (1963/73, 1) uses "correct", Sainsbury (1991/2001, 6) "good" (but also "ought"), and Restall (2006, 1) "good".

[33] Though for concerns with even that assumption see, e.g., Hazlett (2013).

strued in one's logic, they may involve giving our actual words meanings they very clearly do not have. Or they may involve circumstances where no reasoners, including ourselves, exist. If the interest in truth-preservation is parasitic on the value of something like true belief, validity seems in danger of radically overshooting whatever explains the value of good inference.

Though I don't have space to go into details, I think Dogramaci establishes an impressive case for this last conditional. This drives him to embrace an unusual and, I think, problematic substitutional construal of validity owing to Quine (1970/86), since he thinks we can give a story about the value of validity, so-construed.[34] As I say, I won't rehearse Dogramaci's arguments here. The reason, which should by now be clear, is that I think we should reject an underlying assumption of Dogramaci's discussion: that the goodness of an inference stems from the goodness of true *belief*. An intriguing feature of Dogramaci's discussion of why good inference is good is that there is no explicit discussion of what inference is or what it is for. It is seemingly presumed that the only thing inference could be for is to arrive at true beliefs. This *does* make it a mystery why non-actual truth-preservation would be of any value, especially across the extraordinarily broad range of cases validity requires.

To be clear: it is not that I think deductive inference is in no way valuable for sometimes enabling us to arrive at new true or rational beliefs. It is rather that this aspect of its value is a downstream consequence of deductive inference's more general purpose of executing a total information preserving transition over members of the class of acceptance states, of which beliefs form a proper subclass. But if we want to understand what makes an inference good, we have to look to the most general function of inference. Otherwise, we are led precisely into the mystery that animates Dogramaci's investigation.

Once we take the proper perspective, the answer to Dogramci's question of what makes a valid deductive inference good is entirely straightforward. The validity of an inference contributes to its goodness in the sense of its goodness *qua* inference. The goodness of deductive inference requires necessary truth-preservation, because its goodness simply consists in successfully executing a total information-preserving transition. Executing such a transition is its goodness-fixing proper function. And logical validity secures the property of necessary truth-preservation needed for a total information-preserving transi-

[34] Some familiar grounds for concern with the substitutional conception of validity can be extracted from Etchemendy (1990, 2008).

tion to occur. Nothing less would do.

Why does truth-preservation in far-fetched cases that are known to be non-actual matter to valid inference? The question does not even have force except under the presumption that the goodness we are investigating is somehow specially concerned with the actual world. And it isn't: extracting the information from an information bearing state is a process that is completely indifferent to actuality—as is apparent the moment we consider why an inference is good or not when we engage in counterfactual suppositional reasoning. Whether the actual world is compatible with a starting acceptance state or not is simply irrelevant to the issue being settled by deductive inference: what implicit informational commitments does a given representational state have?

Dogramaci's investigation is symptomatic of a more general problem with investigating logic's importance through the states that inference mediates between rather than through inference itself. The problem in this case is that there is no direct route from the value of various acceptance states (especially those like supposition) to that of inference. And this is to be expected. An inference is a mental mechanism operating on acceptance states, much as a hammer is a tool that operates on nails. There is no reason to think that inference derives its goodness-making features immediately from promoting goodness-making features of acceptance states (should there be any such features) any more than there is reason to think that a hammer must derive its goodness-making features from its being usable to fashion good nails.

There are other ways the focus on states can mislead. For example, several of the authors I've discussed—Harman, MacFarlane, and Steinberger—are looking for norms that govern reasoned change in belief. So-construed reasoning is an activity that *takes time*. What is interesting is that, on their face, all the principles we have seen so far are *synchronic* principles: they govern how one's belief states should be organized at a single moment in time.

The natural response, illuminatingly made by Steinberger, is to claim that the synchronic norms can be extended to diachronic ones. Steinberger claims that "A theory of reasoning …is concerned with the dynamic 'psychological events or processes' by which we form, revise or retain beliefs."[35] But he notes that his principle (S), like others found in the literature, is strictly speaking stated in synchronic terms.

(S) If according to S's best estimation at the time, S takes it to be the case

[35] Steinberger (2019a, 307).

> that $A_1, \ldots, A_n \models C$ and S has reasons to consider or considers C, then S has reasons to (believe C, if S believes all of the A_i).

Accordingly, he claims this synchronic principle feeds into a corresponding diachronic norm, one half of which is given by (†-1).[36]

(†-1) If, at time t, S believes that $A_1, \ldots, A_n \models C$ and S considers C or has subjective reasons to consider C, then if S's reasons for believing the A_i are not outweighed by sufficiently strong prior reasons for doubting C, then S has reasons to believe C at t' (where t' is preceded by t).

This principle is complex, but we can ignore some of that complexity as I merely want to zero in on its treatment of time. Focusing on that issue, we can see that the principle as it stands is ambiguous, and it is not clear that any disambiguation is plausible.

For example, is the principle telling us for *any* two times t and t', such that the former precedes the latter, when one has all the reasons and beliefs given at t one has reasons to believe C at t'? Surely not: if years have passed, one's beliefs and one's evidence, so one's reasons, may have dramatically changed. Surely then one may have no reason to believe C. Perhaps we should say that t *immediately* precedes t'. But this seems to presuppose that time is discrete. And, even if it were, wouldn't it need to be a necessary truth for the principle to be sufficiently general? But if time is, or can be, continuous, which times should be linked by our principle? It seems that different persons may be capable of reasoning at different speeds. Should the principle be relativized to the individual's capabilities? Note also that the principle is triggered by having reasons to consider the conclusion C. But what if even though one has reasons to consider C, one hasn't yet? Couldn't it be appropriate to first consider C (given one's reasons for doing so) before coming to believe it? And surely this takes time, at least for many agents. How should this be taken into account?[37]

We could go on. It should be clear by now where these questions are leading. Steinberger starts by formulating a principle of 'reasoning' as a synchronic constraint on attitudes, and then attempts to get back the diachronic element of the principle by considering those attitudes over time. This appears to be

[36] Steinberger (2019a, 322). (†-1) speaks of belief where (S) talks of 'best estimation'—the reason for this disparity won't matter here.

[37] See Podgorski (2017) for related concerns.

getting the methodology backwards. Intuitively reasoning is not merely having one attitude, then having another (and happening to have them align in some way). Rather it is, precisely as Steinberger puts it, an event or process that transitions between such states in a certain way. What we want to evaluate is the event or process—the transition itself. *That* is the thing that is done well or poorly. Principles like (†-1) are trying to get at that event or process in an indirect way, and it is not clear that this is possible. Or at least, it is not clear that it is possible without increasing layers of caveats and conditions, which simply obscure the real target of normative inquiry.

No such concerns afflict a principle like (Good) or (Bad-Appreciation). These principles evaluate a process, event, or activity involved in reasoning directly. There are no concerns that arise from timing, since the timing is 'built into' the object of evaluation, however long it takes, whenever it begins or ends.[38]

The lesson from both Dogramaci and Steinberger is simple. Taking logical normativity to govern states runs into serious troubles, even if we take that form of normativity to be evaluative. The standards that apply to states and acts, and even their metaphysical structure, diverge in significant ways. Investigations into logic's influence that depend on those standards, or those metaphysical features, are going to be distorted to the extent that states are privileged to the exclusion of mental acts.

3.5 The Fallible, the Misguided, and the Obtuse

I want to highlight a final issue that has played an important role in the formulation of bridge principles, and whose relevance to the account of Chapter 2 may need clarification: how epistemic conditions of awareness of logical relations influences logic's normative grip. The principle (Good) that I advanced in §3.2, for example, integrates such an epistemic condition—the 'appreciation' of the necessarily truth-preserving character of an inference. What

[38] Obviously, this section reveals my sympathies with the idea that inference is evaluable diachronically. For some good reasons to think this, see Hlobil (2015), who deepens considerations one also finds in Broome (2013, 2015). I do, however, acknowledge grounds for worrying that this is mistaken. See, e.g., Valaris (2017), building on concerns going back to Strawson (2003). Fortunately, the view of Chapter 2 does not prejudge this issue. Relatedly, a feature of (Good) is that it does not *require* norms governing inference to diachronic. It is merely that if our final understanding of inference is as diachronic in nature, (Good) will apply to it in those terms.

is important, but may not initially be clear, is that this epistemic condition is playing a very different role from the roles of other epistemic constraints in existing bridge principles. Unlike with many of the other features of (Good) that I've discussed so far, I don't think this presents a special *advantage* enjoyed by (Good). Rather, it is just a feature of the very different approach to logical normativity that I'm advancing.

To get clear on all this, it is worth reviewing how the issue of epistemic constraints arises in the project of formulating bridge principles. That project runs up against the issue of epistemic constraints in the form of a dilemma, brought on by the contrast between the *logically obtuse* on the one hand, and the *merely fallible* or the *rationally misguided* on the other.

The logically obtuse are those who fail to see even the most basic entailment relations. Reflection on their existence seems to show that some rational norms connected with logic apply to individuals regardless of whether they recognize that a logical entailment holds. Consider a person who believes the major and minor premise of an instance of Modus Ponens, with excellent evidence, and who is considering its conclusion, which they have no evidence against. It would seem irrational for such an individual not to draw the conclusion even if they can't muster the ability to recognize the validity of Modus Ponens—for that inability of itself renders them guilty of irrationality anyway. It accordingly seems desirable to have logical norms that succeed in condeming this logically obtuse individual.

The logically fallible are those who fail to see entailment relations of higher complexity, and the misguided are those who are led by attractive reasoning to incorrect beliefs about logic. These categories comprise a much larger group. We all reasonably fail to recognize logical entailment patterns of sufficiently high degrees of complexity. And sometimes we may be reasonably led, perhaps by what appear to be good arguments, to think a pattern constitutes a logical entailment when it in fact does not. We intuitively would not like logical norms to impugn those former individuals who fail to infer remote consequences of their current attitudes. And perhaps we would even want logical norms to excuse at least some of the latter cases of the logically misguided as well.

The obtuse thus seem to call for logical norms indifferent to an agent's logical awareness, whereas the fallible and misguided seem to call for norms with epistemic caveats. Defenders of bridge principles have tried to negotiate this tension in various ways, all of which seem unsatisfactory.

For example, MacFarlane sought to avoid epistemic triggering conditions in his bridge principles, emphasizing that ignorance of logic should not generally be exculpating.[39] If we build strong epistemic triggering conditions into our norms, "[t]he more ignorant we are of what follows logically from what, the freer we are to believe whatever we please—however logically incoherent it is. But this looks backwards. We seek logical knowledge so that we will know how we ought to revise our beliefs: not just how we *will* be obligated to revise them when we acquire this logical knowledge, but how we are obligated to revise them even now, in our state of ignorance."

He accordingly navigated the tension between the obtuse and the fallible by appealing to a pair of principles. His stronger principle, (wo-), merely forbade inconsistent attitudes.

(wo-) If $A, B \models C$ then you ought to see to it that if you believe A and you believe B, you do not disbelieve C.

This allows the fallible to escape rational criticism by remaining agnostic. A weaker principle, (wr+), gave reasons for positive attitudes, but did not obligate one to have them.

(wr+) If $A, B \models C$ then you have reason to see to it that if you believe A and you believe B, you believe C.

This allows us to say that the obtuse have *some* reasons to draw the conclusions they are failing to draw.

For the reasons I've already mentioned in discussing Steinberger, I feel that (wr+) is an exceedingly weak and non-committal principle because it says nothing about the strength of logical reasons. We need a specification of the source of this strength, and an understanding of it. Without those things, it is not clear we understand why the the willfully obtuse are irrational (for example, perhaps the obtuse don't *feel* like they can be bothered to draw the conclusion—is that not *some* reason of some strength not to draw it? (wr+) on its own seems helpless to tell us why any such reasons aren't defeating).

And despite merely forbidding, rather than requiring, attitudes, (wo-) still becomes too strong in the absence of epistemic triggering conditions. Field

[39] Well, at least he avoided epistemic triggering conditions governing validity itself. He eventually suggests that recognition of the schematic form that an inference takes may be needed as a triggering condition—see below.

(2009, 254) supplies an illustrative case of the merely fallible: "…it is natural to suppose that *any* rational person would have believed it impossible to construct a continuous function mapping the unit interval onto the unit square, until Peano came up with a remarkable demonstration of how to do it. The belief that no such function could exist (in the context of certain set-theoretic background beliefs) was eminently rational, but inconsistent."

Field (2009) took the opposite extreme from MacFarlane, simply building in that logical relations should be 'obvious' in order for logically governed epistemic norms to have application. For reasons I'll discuss shortly, I applaud this maneuver, and will ultimately take up a version of it myself. But we need to be careful about what the notion of 'obviousness' is, and how it is to be applied. As I discussed in §3.1, Field's obviousness constraint interacts in a problematic way with the kinds of norms he introduces in his bridge principle. This contributes to the sense that he effectively exploring norms for believing obvious truths, logic just being one way such obvious truths are supplied.

In a later paper, Field seems to back off of his obviousness constraint. He says instead (crediting the idea to MacFarlane):

> To handle [the fallible and the misguided] …we recognize multiple constraints on belief, which operate on different levels and may be impossible to simultaneously satisfy. When we are convinced that a certain proof from premises is valid, we think that in some "non-subjective" sense another person *should* either fully believe the conclusion or fail to fully believe all the premises—even if we know that he doesn't recognize its validity (either because he's unaware of the proof or because he mistakenly rejects some of its principles).[40]

The suggestion is that logical norms are non-subjective "in some sense."[41] It is not clear that in his later paper Field has a specific interest in capturing norms for reasoning broadly construed. But it is worth noting that there are serious obstacles for an 'objective' norm to play such a role.

[40] Field (2015, 44).

[41] I noted concerns for the idea that logical norms are objective in discussing MacFarlane's treatment of the preface paradox in §3.1. But those concerns don't apply to Field here. The problem for MacFarlane's case was that his objective norms would seemingly have to concern the availability of more *a posteriori* evidence, which was an implausible role for logical norms to take on. But that is not a concern for Field's move, since his more objective norms can plausibly concern only additional *logical* knowledge or awareness.

As I understand the proposal, Field's objective norm avoids pronouncing (some) fallible and misguided reasoners subjectively irrational because they simply lack the logical knowledge or awareness that we have. The best sense I can make of the idea that the fallible nonetheless 'objectively ought' to change their attitudes to conform with the principle is that they would have to conform to it, if only they had the knowledge we did. What then are we to say about the obtuse? They too seem to lack logical knowledge of elementary logical relations. So the merely objective norm would intuitively fail to pronounce them subjectively irrational as well—at least until we add some further conditions or caveats.

The basic idea here is that there is very little difference between a 'subjective' norm which has as a triggering condition that agents know relevant validities, and an 'objective' norm that applies to agents only in the sense that 'if they knew better' this is how they should reason. Both intuitively fail to say anything about the obtuse, and for the same reason: the norms have their 'real force' only against the assumption that the agent knows or otherwise has information about validities. Accordingly the same concern for subjective norms with strong epistemic triggering conditions seems to arise for objective norms that Field later advances—at least if we are concerned with norms for reasoning broadly construed.

Steinberger, it may be worth recalling from §3.1, introduces one of the weakest epistemic triggering conditions into his bridge principles (that a logical entailment hold "according to the reasoner's best estimation"). He motivates this by saying that if bridge principles are giving directive norms, then it is plausible to have weak triggering conditions to ensure the norms can be followed. My objection to this was that being 'followable' doesn't require a condition quite as strong as the one Steinberger appeals to. Certain logical consequences are not only *a priori*, but incredibly obvious. Anyone can in-principle follow them. Allowing agents with completely perverse logical attitudes to escape from even these basic requirements is no more plausible in a directive norm than in any other kind.

So far I've argued that neither MacFarlane, nor Field, nor Steinberger successfully addresses the question of how to jointly handle the fallible or the misguided and the obtuse. Can the view of Chapter 2 contribute to our understanding of this situation? Partially. Let me begin with a claim that I think should be appreciated independently of my account of logical normativity, be-

fore turning back to what is distinctive of that account.

One of the most vexing aspects of the tension in accounting for the obtuse and the fallible is that we intuitively draw a line between two kinds of logical relations: those which one is rationally accountable for seeing, and those that one is not. The obtuse make mistakes with respect to the former principles. The merely fallible and misguided make mistakes with respect to the latter. The line may be fuzzy and it may be context-sensitive. But what seems clear is that we draw it.

So if we want to account for the difference between the obtuse and the more reasonable agents, that line between two types of entailments must be drawn somehow. The line is clearly not the line between the entailments known to an agent and those not known. Nor is it the line between the entailments believed by the agent, and those not yet believed by them. The line is more objective than that. That is why one can't get always get around the grip of logical norms by simply failing to believe important consequence relations.

The first thing I would like to suggest is that reflection on this matter should lead us to believe that the line is a *normative* one. Intuitively, we are out to find a distinction between those entailments one is *rationally required* to see, and those one is *rationally permitted* not to see. I do not mean to presuppose that it is 'fundamentally normative.' Perhaps it can be cashed out in non-normative terms. But we are interested in the line we are drawing because of its direct ties to what ought to be recognized.

The second thing I want to suggest is that, whatever one's interest in the normativity of logic, this task should be 'outsourced.' That is, the question of which logical truths we are required to see is not a question that logic, or a view of the normative force of logic, should itself be settling. It belongs to a more general sphere of epistemic requirements that has nothing special to do with logic.

The recommendation I am making here is similar to one MacFarlane makes in discussing the 'recognition of logical form.' MacFarlane eventually considers building in a kind of epistemic triggering condition into his bridge principles. But, curiously, the trigger does not concern awareness of an inference as valid. Rather it concerns awareness of the logical form of a potential inference. MacFarlane suggests that one may have to see logical form correctly for logical norms to start applying. For example, one may need to recognize that a single name is being used over and over in several premises in order to

know what one can infer from those premises.

We needn't delve too deeply into this idea, or the justification for it, here. The important thing is that MacFarlane goes on to note that the bridge principle he endorses will not condemn a certain kind of 'formally obtuse' reasoner—the person who doesn't recognize the logical forms that they should. Here is his discussion and response:

> It might be objected that this account restricts the application of logical norms too far. Shouldn't we sometimes be held accountable for failing to apprehend logical structure that really is there, or for taking there to be logical structure that isn't there? Sure. But I am inclined to keep these norms *for* apprehension separate from the logical norms that arise *from* the apprehension of inferences as instances of formally valid schemata. The former seem to group together with general epistemic norms enjoining careful observation and thorough investigation, not with specifically logical norms.
>
> MacFarlane (ms/2004)

It may be that one *should* see certain logical forms. But it is not logic that instructs you to see them. As MacFarlane puts it, more general epistemic norms of care will do that work. The 'formally obtuse' individual who fails to draw an obvious inference may be failing to see the obviously logical form of the inference they could make, or they may see that form and fail to draw the inference. Either way, they are irrational. But their irrationality is of a different kind, depending on which mistake they make. In particular, the former mistake is not clearly a logical one.

I would put the general point here by saying that a norm which *concerns* logic is not necessarily a norm *of* logic. I applaud MacFarlane's use of that distinction. I only question why he didn't extend this treatment to awareness of validities. I agree that if awareness of logical form is part of responding correctly to logical norms, it need not be logic that pronounces on which of them we 'ought' to see. But why not also say that although awareness of entailment relations is part of responding correctly to logical norms, it need not be logic that pronounces on which of *them* we 'ought' to see?

This, I think, is the intuitively correct diagnosis of what happens with the 'ordinary' logically obtuse with which we began this section. As with MacFarlane's formally obtuse just discussed, the ordinary obtuse can be making

several different *kinds* of mistake. They may be willing to draw any logically valid inference they can (given further appropriate epistemic conditions), but fail to see an inference is valid. Or they may see an inference is valid, but be unwilling to draw it. In the former case, the obtuse are not obviously making a kind of mistake that logic specially condemns: they simply fail to see something obvious—general epistemic norms of care should account for that. In the latter case, however, we come closer to flouting properly logical dictates.

This is why I applaud Field's early approach to the problem of the obtuse and the fallible. He simply labels certain inferences 'obvious' and uses this to trigger the application of logical norms. The notion of obviousness goes largely unexplained. Field even suggests that the reference to the 'obvious' logical entailments could be dispensed with in favor of simple listing of those that mattered. As should be clear, this doesn't solve or otherwise illuminate the problem of the obtuse. Rather it outsources the problem, as should be done, since it doesn't properly concern the dictates of logic themselves.

Once this move is made, though, we can start to see the importance of Chapter 2 come into view. The problem is that even once we have outsourced this work for awareness of logical principles to play, we have *not exhausted* the role of some such awareness in formulating logic's normative force.

To see this, let us consider one of the simple, 'obvious' consequences that any reasoner should recognize—let's say Modus Ponens. An obtuse individual, Donald, believes p and believes that if p then q (with good evidence), but does not recognize that it follows from this that q (where the question of whether q is pressing, he has no evidence against q, etc.). But he somehow comes to believe q anyway. He is, of course, unable to *see* why q is true. He simply believes it.[42] Was this a way for Donald to properly conform to the dictates of logic? I claim that this is not clear.

It may not be easy to see the concern here, precisely because it is hard to imagine someone in Donald's predicament. It may accordingly be useful to contrast a case where someone does what Donald has done with an entailment that is more easily understood as unappreciable. Suppose Bill, an ordinary undergraduate, has just learned the axioms of Set Theory (and associated definitions of number, etc.). He comes to believe that there is a continuous function

[42] We may even add that believes q 'on the basis' of the the fact that p and the fact that if p then q, as long as this is compatible with his failing to see the connection between the premises and the conclusion.

mapping the unit interval onto the unit square. Bill may even base that latter belief on the set theoretic axioms and relevant definitions. But he acknowledges that he can't see why the function exists merely on the basis of those axioms. Indeed, he can't say a word about why the connection between these claims holds.[43]

What Bill has done is performed a kind of inferential leap in the dark. Donald has done this too, though it may be less obvious, precisely because of how obvious we find the entailment he cannot grasp. Something has gone wrong with both. But what?

It's not that Bill, or Donald, has done something logic forbids. How could they? Logic informs us only that what they came to believe followed logically from their premises. Nor have they failed to do something logic (on its own) specifically requires. The fact that a logical entailment holds does not require one to make inferences involving it. This is true even for Donald—that was part of the lesson about outsourcing we drew from the logically obtuse.

The answer is that these characters have done something poorly. In particular, they have poorly performed the very activity whose goodness logic helps track. They have inferred (or at least, our story is consistent with them having inferred), but they have failed to appreciate what makes their inference necessarily preserve truth. That form of appreciation, as noted in Chapter 2, is customarily viewed as a condition on an inference's being performed correctly. It is not enough to make an inferential transition in a way that necessarily preserves truth. One must also appreciate that this is so.

This is why (Good) is formulated as it is.

(Good) If $A, B \models C$, any inference from A and B to C, in which the inference's necessary truth-preserving character is appreciated by the inferrer, is good *qua* deductive inference.

Conditions on which entailments should be recognized by inferrers (essentially what theorists appealing to epistemic triggering conditions have been aiming to capture) are not stated anywhere in (Good). This is appropriate: for the reasons I've been giving, those conditions really have nothing special to do with logic, and belong to the more general sphere of epistemic reasonable-

[43] Cf. cases of 'large leaps' in inference involving Fermat's Last Theorem discussed in Boghossian (2003), Berry (2013), Dogramaci (2015a), and Schechter (2019). I will discuss these kinds of cases in somewhat more detail in Chapter 5.

ness and caution. But there is a separate way in which a certain kind of knowledge should figure in a principle explaining logic's normative force. Logic only ever supplies us with a necessary condition on the goodness of inference. As such, even with respect to that limited evaluative form of normativity, logic in a way tells us nothing on its own. We must pair the conditions on content uncovered by logic with a separate, but connected, necessary condition to arrive at a reasonable normative principle. It's not so much that this added necessary epistemic condition is 'required by logic.' Indeed, the thing that must be appreciated is not itself formulated in logical terms. It's that we cannot fully understand the kind of goodness logic investigates except by reference to that epistemic condition. Only in conjunction with it do the conditions tracked by logic have any normative implications for reasoning at all.

3.6 Taking Stock

I've claimed that logic's normative import is most fundamentally directed at evaluating inferential acts and that, once this is seen, the normative force in question is simple and exceptionless. No element of this view is new. Logic texts often cash out the normative implications of validity in evaluative terms, and we have an extensive literature explaining the function of evaluative talk in precisely the ways I recommend. What is more, the default view, including the view of most authors I've discussed, is that logic is relevant to reasoning which is itself a process, or activity, or event, and not merely a state or group of states. And there is of course an extensive and growing literature on inference itself, in which inference is customarily viewed as just such an activity, process, or event.

So why have investigations of logic's normativity seemingly ignored this perspective? My suspicion is that the trajectory of the literature owes to a single, simple, and understandable move of Harman's. In asking how logic could be relevant to reasoning, Harman formulated his principles in terms of constraints on attitude states. This was a reasonable thing to do, given that Harman was partly moved by the attractions of a view of logic on which it merely supplies us with general truths, and a view of inference on which there was very little logically distinctive of it as a process (see n.22).

It was certainly reasonable to explore this as one avenue among many. But it was ultimately a mistake. Logic doesn't merely supply us with general truths,

but necessary truth-preserving relations among contents. And it does this precisely because this is a necessary condition on good deductive inference. To the extent logic has distinctive normative implications for reasoning, they apply at the level of inferential transitions.

But once Harman's reasonable move was made, it kicked off an equally reasonable research program: that of finding bridge principles capturing logic's influence on acceptable combinations of beliefs or credal states. This program is not merely reasonable because of the reasonableness of Harman's starting point. Rather, it is because there is no reason I know of (including no reason from anything I have said in this Chapter) to suspect that this program cannot reach an end. There is no reason to think that we cannot formulate a bridge principle of the sort that MacFarlane catalogues which, encrusted with enough caveats and conditions, is true. Indeed, I see the current literature as making steady progress toward that very goal.

So it is important to note that my objection to this program is not that it is somehow unfulfillable. It is rather that it has nothing to do with what is distinctive of logic and its normative relevance to reasoning. Instead, the program has much more to do with a great admixture of epistemic norms, including especially norms governing reasonable formation and management of beliefs in response to evidence and recognized truths, to which logic may contribute partially and indirectly (mostly by supplying us with certain truths, or organizing our evidence). One thing this means is that finding the true bridge principles will probably be a convoluted and frustrating task, with the resulting principles becoming increasingly qualified, precisely as we have seen. But more importantly, even if this task reaches an end, there will be no way to 'factor out' of our final principles what was distinctively contributed by logic. By the end, logic's role will be completely swamped by the complementary, overlapping, and interacting non-logical epistemic norms.

This concern is most easily seen from the fact that standard bridge principles, out of the starting gate, seemingly give up on saying why the resources of logic condemn logically fallacious reasoning. As we saw, no bridge principle seems to come close to explaining what is wrong with affirming the consequent. By now it should be clear why this is so. What is wrong with the person who has engaged in such a fallacy has nothing to do with the combination of their attitudes, even over time. What is wrong with them is the process or event that led them to those attitudes. That simple idea was understandably set aside

with Harman. But to make any headway in understanding logic's normative force, we must take it up again.

CHAPTER 4

The Impossible and the Unthinkable

We've just seen in Chapter 3 how even a skeletal account of inference and inferential goodness can transform our understanding of debates in the foundations of logic. In this chapter and the next, I'll argue that an issue historically framed in logical terms can conversely give us insight into the nature and structure of deductive inference.

The issue I'm alluding to arises from a tension between two ways that philosophers have viewed the role of logic in constraining possible thought. The first way of thinking comes from a storied tradition in the philosophy of mind leading from the ancients through the medievals and moderns to the early analytics. On this view, the impossible—and especially the logically impossible—is literally unthinkable. Alongside this tradition, however, there is a second, diametrically opposed way of viewing logic on which illogical thought is taken to be a familiar and pervasive feature of fallible human cognition.

My goal in this chapter is to extract some minimal, compatible insights from these opposed and controversial perspectives, and use them to reveal a hopefully less controversial, but still puzzling, feature of the space of possible thoughts. The puzzling feature is that while most propositions become easier to entertain as their complexity is reduced, there is a special subclass of propositions (roughly, the impossible ones) that appear to become *harder* to entertain as their complexity is reduced.[1] I provide some limited empirical support for

[1] Here and elsewhere, I use "entertain" as a blanket term to cover different ways of taking attitudes to propositions. On this usage, to believe a proposition is one way of entertaining it, and suppose it is another way to entertain it, and so on. Some philosophers think there is a distinctive way of 'grasping' a proposition that can (or must) precede any attitude being taken toward it. If there is such a way of grasping a proposition, it should probably also constitute a form of entertaining on my usage.

this duality in §4.2. Then I suggest that we best explain the duality by positing a demanding cognitive relation of *representational crowding-out* which, when borne by an agent to a proposition, precludes that agent from entertaining that proposition. I develop a theory on which this relation arises as a *mode* of propositional representation, and explore some features that the relation would bear on that theory.

Eventually in Chapter 5, I'll argue that the cognitive relation of crowding-out is a key missing ingredient from accounts of deductive inference. But we will need to get much clearer on the what this relation is before we can understand how it can play that role. So let's turn to examine the relation first.

4.1 Two Perspectives on the Impossible and the Unthinkable

I want to begin with a short historical preamble, to give a flavor for the opposing perspectives on the impossible and the unthinkable just alluded to.[2] I will be quick, and will hardly be able to do justice to nuances in the positions of the various authors I mention. This will hopefully not matter much: my principal aim is merely to bring out recurring general motivations and examples to discuss in §4.2.

In the *Metaphysics*, Aristotle lays out founding arguments for his version of the law of non-contradiction—the claim that "the same attribute cannot at the same time belong and not belong to the same subject in the same respect." (*Metaphys.* IV 3 1005b19–20)[3] Alongside these considerations Aristotle also argued for what is sometimes called the 'psychological law of non-contradiction': "It is impossible for any one to believe the same thing to be and not to be" (*Metaphys.* IV 3 1005b24). Aristotle's arguments for the psychological law of non-contradiction are notoriously slippery.[4] But for our purposes, Aristotle's explanations of the psychological law are less important than the fact that he took it to be a law at all—something *in need of explanation* just as much as the law of non-contradiction itself.

The suggestion that logic may constrain not only reality but how we think of it seeped into scholastic thought. Aquinas made use of this idea in trying to explain how we avoid infringing on God's omnipotence in saying that God

[2] The discussion here owes much to, and at many places closely follows, that of Conant (1991).

[3] I use the translation in Aristotle (1991).

[4] See, e.g., Łukasiewicz (1910/1979) for critical discussion.

lacks the ability to violate the laws of logic.

> . . . it is better to say that such [impossible] things cannot be done, than that God cannot do them. Nor is this contrary to the word of the angel, saying: "No word shall be impossible with God." [(Luke i.37)] For whatever implies a contradiction cannot be a word, because no intellect can possibly conceive such a thing."
>
> (*Summa Theologica*, Q. 25, Art. 3.)[5]

As stressed by Conant (1991), Aquinas's final appeal to the unintelligibility of contradictions seems to play an important role in his defense of God's omnipotence. Aquinas is attentive to the danger of saying that there are some conceivable scenarios—the impossible ones—that God cannot actualize. To say this seems to bring with it the intelligibility of acts that bring about those impossible scenarios, with only some of those intelligible acts being possible for God. If so, God could not do everything conceivable. We can avoid this conclusion if we maintain that contradictions "cannot be a word"—that is, do not express something that could be thought, even by God. In this way, God's power can extend without exception to every conceivable action, including the creation of every conceivable reality.

Descartes followed Aquinas in taking the limits of the possible to likewise delimit our cognition, but denied that logic likewise bounded divine intellection or power. On Descartes' view, the things we view as necessary truths are indeed necessary, but only because contingently willed so by God. He expresses this in a letter to Mesland May 2, 1644:

> [E]ven if God has willed that some truths should be necessary, this does not mean that he willed them necessarily; for it is one thing to will that they be necessary, and quite another to will this necessarily, or to be necessitated to will it.
>
> (Descartes, 1991, 235)

God could have brought about what is in fact impossible. But He chose not to do this, and instead chose to make those things impossible. And at the same time, He furnished us with minds whose powers of conception were bounded by the very things He willed impossible. From a letter to Arnauld, July 29, 1648:

[5] Using the translation in Aquinas (1981).

> I do not think we should ever say of anything that it cannot be brought about by God. For since every basis of truth …depends on his omnipotence, I would not dare to say that God cannot make a mountain without a valley, or bring it about that 1 and 2 are not 3. I merely say that he has given me such a mind that I cannot conceive a mountain without a valley, or a sum of 1 and 2 which is not 3; such things involve a contradiction in my conception.
>
> (Descartes, 1991, 358–9)

Spinoza, while maintaining a radically different modal metaphysics from Descartes, nonetheless shared with him the view that contradictions cannot so much as be entertained or imagined. This comes out in Spinoza's discussion of Chimaera, which are things "whose nature implies that it would be contradictory for it to exist" (C I 24), like a round square.[6] Spinoza says that neither the intellect nor the imagination (which for Spinoza exhaust the cognitive) can accommodate Chimaera, so that they are 'merely verbal':

> [I]t should be noted that we may properly call a Chimaera a verbal being because it is neither in the intellect nor in the imagination. For it cannot be expressed except in words. E.g., we can, indeed, express a square Circle in words, but we cannot imagine it in any way, much less understand it.
>
> (C I 307)

And in the *Treatise on the Emendation of the Intellect* he elaborates:

> …we say "Let us suppose that this burning candle is not now burning …". Things like this are sometimes supposed …But when this happens, nothing at all is feigned.
>
> (C I 26)

When we purport to suppose a contradiction, nothing is 'feigned'—nothing is entertained or imagined—so that the would-be supposition is at last understood to be impossible, a kind of mock-supposing.

[6] "C I" citations are to page numbers in volume I of Curley's translations in de Spinoza (1985).

All this makes sense of how Hume, in his *Treatise*, was able to claim that it was an established fact that what is clearly conceivable cannot include what is "absolutely impossible"—which for Hume corresponds roughly to the logically impossible.

> 'Tis an establish'd maxim in metaphysics, *That whatever the mind clearly conceives includes the idea of possible existence, or in other words, that nothing we imagine is absolutely impossible* ... We can form no idea of a mountain without a valley, and therefore regard it as impossible.
>
> (*A Treatise of Human Nature* 1.2.2.8)

Kant, on some (admittedly controversial) readings, not only took logic to delimit possible thought but to be, definitionally, a study of that delimitation.[7] The idea appears to be brought out in passages like the following in the *Jäsche Logic*.

> [The] science of the necessary laws of the understanding and of reason in general, or what is one and the same, of the mere form of thought as such, we call *logic*.
>
> As a science that deals with all thought in general, without regard to objects as the matter of thought, logic ... [is to be regarded as] a science of the necessary laws of thought, *without which no use of the understanding or of reason takes place at all* ...
>
> (*Jäsche Logic*, 14-5, my emphasis)[8]

Here Kant has sometimes been read as saying that there is no use of the understanding—no thought—whatsoever that fails to be in conformity with logic's dictates.

The inheritor of this perspective in the early analytic tradition is (the early) Wittgenstein. He casts the importance of the *Tractatus* as follows:

[7] These broadly 'constitutivist' readings like those in Conant (1991), Putnam (1994), Tolley (2006), and Merritt (2015) contrast with 'normativist' readings on which logic for Kant merely rationally requires logical thought. For examples of the latter readings, see MacFarlane (2002), Longuenesse (2005), Anderson (2005), Hanna (2006), Stang (2014), Leech (2015), and Lu-Adler (2017). For a view that tries to reconcile both traditions, see Nunez (2018).

[8] I use the translation of Kant (1992).

> The book will therefore draw a limit to thought, or rather—not to thought, but to the expression of thoughts …
>
> (Wittgenstein, 1922, 3)

It is familiarly contested precisely how Wittgenstein meant to do this. But whether Wittgenstein is properly understood as having drawn, or 'shown', or intimated, or dissolved the limit of which he speaks in his preface, what seems clear is that he does it somehow with the help of logic. And many passages in the *Tractatus* seem to bear out an interpretation on which Wittgenstein takes the illogical to be unthinkable. For example:

> That logic is a priori consists in the fact that we *cannot* think illogically.
>
> (Wittgenstein, 1922, §5.4731)

Roughly this idea reappears yet again in the work of Husserl. In his *Formal and Transcendental Logic*, Husserl distinguishes between two types of judgments: the non-explicit and the explicit. It is clear that 'explicit' is a kind of honorific or ideal of judgment (a 'proper' form of judgment). But as regards such explicit judgment, Husserl maintains that "only one [member of a contradictory pair] can be accepted by any judger whatever in a proper or distinct unitary judging."[9] He further maintains that once a thinker sees a consequence of a proper judgment she "not only judges the consequence in fact but "*cannot do otherwise*" than judge it."[10]

Echoes of the idea that logic delimits the scope of the intelligible continue to be found in contemporary work, both in a philosophical tradition that takes up the mantle of the foregoing historical positions,[11] but also among working logicians. Indeed, on the opening page of the widely used introductory text, *Logic, Proof, and Language*, we find the following as an explanation of logic's fundamentality.

> …there is an overwhelming intuition that the laws of logic are somehow more fundamental, less subject to repeal, than the laws

[9] Husserl (1969, 190).

[10] Husserl (1969, 189). It is worth noting this is appears to be a reversal of Husserl's earlier position in his *Logical Investigations*. Thanks to Manish Oza for pointing me to these passages.

[11] See, e.g., Putnam (1994), Kimhi (2018), Travis (2019), Marcus (2020, 2021), Oza (2020).

> of the land, or even the laws of physics. We can imagine a country in which a red traffic light means *go*, and a world on which water flows up hill. But we can't even imagine a world in which there both are and are not nine planets.
>
> (Barwise & Etchemendy, 1999, 1)

There are, of course, significant differences between all the philosophers above, not to mention very challenging questions about the correct interpretation of the remarks I've cited. But in spite of these differences and nuances, it seems reasonable to say that the broad idea that the impossible, and especially the logically impossible, has some role in limiting cognition is a common thread among them. The view's recurrence reveals its enduring philosophical appeal within quite different philosophical frameworks.

What is striking is that alongside this tradition one appears to find diametrically opposed considerations, presented less as a rival philosophical doctrine and more as a simple set of indisputable facts.[12] Lewis, in arguing for the utility of the framework of 'fragmentation' to model corpuses with inconsistent sets of information, casually gives himself—that is, his belief state—as an example of a 'mildly inconsistent' corpus.

> I used to think that Nassau Street ran roughly east-west; that the railroad nearby ran roughly north-south; and that the two were roughly parallel. (By "roughly" I mean "to within 20°".) …each sentence in an inconsistent triple was true according to my beliefs…
>
> (Lewis, 1982, 436)

Lewis, I take it, means for this encounter with inconsistent beliefs to be a familiar phenomenon to his readers, offering no special justification for its possibility.

Not only does it seem possible to have illogical beliefs, sometimes this even seems *rational*. Examples of logical inconsistencies arising in complex mathematical cases seem especially compelling in this regard. Consider a case of Field's, which we discussed in a different context in Chapter 3:

[12] There is, of course, *also* a rival doctrine about the role of logic: normativism, which holds that the laws of logic do not describe how we must think, but rather prescribe how we ought to (which seems to require the possibility of failure). Indeed, some of the authors considered in this preamble have been read as normativists—see, for example, n.7 for a list of Kantian normativists.

> …it is natural to suppose that *any* rational person would have believed it impossible to construct a continuous function mapping the unit interval onto the unit square, until Peano came up with a remarkable demonstration of how to do it. The belief that no such function could exist (in the context of certain set-theoretic background beliefs) was eminently rational, but [logically] inconsistent.
>
> (Field, 2009, 254)

These examples involve belief, as opposed to supposition or imagination, which are the typical attitudes wielded by detractors of illogical thought to emphasize the extent and stringency of logic's constraints on cognition. But, using a similar style of mathematical example to Field, Lewis claims it is possible to imagine the impossible in the very sense we can imagine many other complex scenario.

> It is impossible to construct a regular polygon of nineteen sides with ruler and compass; it is possible but very complicated to construct one of seventeen sides. In whatever sense I can imagine the possible construction, I can imagine the impossible construction just as well. I do not imagine it arc by arc and line by line, just as I don't imagine the speckled hen speckle by speckle—which is how I fail to notice the impossibility.
>
> (Lewis, 1986, 90)

Lewis might have plausibly added: in whatever sense one can suppose or believe the possible construction has been made, one can suppose or believe the impossible one has been made as well.

Graham Priest, expressing incredulity at the passage from Hume quoted above, says that it seems to him that he "can conceive of and imagine *anything* that can be described in terms that I understand"[13] and gives several examples, including the following mathematical one.

> I have no difficulty in conceiving [Goldbach's conjecture], and no trouble conceiving its negation, though one of these is mathematically impossible. Indeed, mathematicians must be able to conceive these things, so that they understand what it is of which

[13] Priest (2016, 2659).

> they are looking for a proof, or so that they can infer things from them, in an attempted *reductio* proof. Nor does the conceivability of Goldbach's conjecture and its negation disappear if I discover which of them is true, and so the other no longer appears mathematically possible to me. Hence, when something is conceived it may not even *appear* to be possible.
>
> (Priest, 2016, 2658)

These remarks come from more theoretically motivated discussions, in which various kinds of illogical thought are probed as phenomena of theoretical interest. But there are many more remarks in a similar vein, in which the existence of logically incoherent attitudes is treated merely in passing. For example, it is not uncommon to hear that one of the primary benefits of philosophical study is to root out contradictions in one's own belief system, and begin to repair them. In the context of critical thinking courses, students often study 'logical fallacies'—which are meant to be very common forms of illogical reasoning that philosophical reflection can help us identify and avoid. But if there is no such thing as illogical thought, framing the value of philosophy and logic in these terms seems to make little sense.

So we appear to have two very different modes of thought about the impossible in its relation to possible thought. On the one hand, we have a tradition spanning millennia that treats certain kinds of thoughts about the impossible, especially certain kinds of illogical thought, as themselves impossible. On the other hand, we have a casual treatment of illogical thought—that is, thought about what is impossible by the lights of any alethic modality—as a simple and commonplace feature of our fallible cognition. Perhaps the existence of such thought is lamentable, and to be minimized, but it is nonetheless pervasive.

4.2 A Duality in the Space of Thought

I've suggestively juxtaposed the foregoing views about the role of the impossible in possible cognition as if they were in tension. Are they? There is no simple answer to that question, even if we focus on particular authors from among those I've mentioned. But I want to set historical and exegetical questions aside here. My goal has mainly been to survey the precedents for a collection of plausible claims I would like to defend, some of which seem to act as ground-level motivations for each of the two modes of thought I've surveyed. What I want

to argue is that this collection of claims taken together is both highly defensible, and reveals an unusual division in the space of possible thoughts that cries out for some kind of explanation.

Before setting up my claims, I need to make one remark about the scope of my inquiry. At the outset I will restrict my attention to single attitudes (e.g., single beliefs, suppositions, or imaginings) that are borne to a single, though perhaps complex, content. This will exclude consideration of multiple attitudes that stand in conceptually or logically contradictory relations. The reason for this focus is that an increasingly popular treatment of such collections of inconsistent attitudes is through the framework of fragmentation, alluded to in the discussion of Lewis above.

On that view, a belief state or other attitude state can be viewed as divided into parts which may come into logical conflict, but which are individually coherent. This helps explain the sense in which it sometimes seems reasonable to think that not everything becomes true according to a collection of inconsistent attitudes. Here is Lewis elaborating on his inconsistent views about streets and railroads:

> …each sentence in an inconsistent triple was true according to my beliefs, but not everything was true according to my beliefs. Now, what about the blatantly inconsistent conjunction of the three sentences? I say that it was not true according to my beliefs. My system of beliefs was broken into (overlapping) fragments. Different fragments came into action in different situations, and the whole system of beliefs never manifested itself all at once. The first and second sentences in the inconsistent triple belonged to—were true according to—different fragments; the third belonged to both. The inconsistent conjunction of all three did not belong to, was in no way implied by, and was not true according to, any one fragment. That is why it was not true according to my system of beliefs taken as a whole. Once the fragmentation was healed, straightway my beliefs changed: now I think that Nassau Street and the railroad both run roughly northeast-southwest.
>
> (Lewis, 1982, 436)

The view I eventually endorse about single inconsistent attitudes will extend to multiple inconsistent attitudes. But it may end up having to interact with techniques like fragmentation, which complicates matters. In part

to simplify, I will start by focusing on the single attitude case. Note that, of course, the strategy of explaining contradictory attitudes through fragmentation does not obviously extend in a natural way to apply to a single-attitude case. The whole idea behind fragmentation requires attitude states to break up into parts with their own coherent batches of content. This fragmentation doesn't obviously—at least without substantial elaboration—make sense in application to a single attitude, borne to a single inconsistent content.

With that caveat out of the way, I want to defend a pair of claims, guided in part by insights of the authors mentioned in §4.1.

(I) For any alethic modality M, there are some M-impossible contents, such that there are possible single attitudes borne to those contents.

(II) For some but not all alethic modalities M, there are M-impossible contents such that:

- there is a characteristic strong resistance to the formation of single attitudes borne to those contents; and
- the resistance in question tends to *increase* as the complexity those contents *decreases*.

If (I) is true, then it is a mistake to think that any form of impossibility, even logical, constrains possible cognition in and of itself.[14] But if (II) is true, then there are contents impossible relative to a particular restricted class of modalities which tend to create a resistance to cognition that is influenced by matters of complexity. I will say more about why these claims would call out for explanation. But first, let me defend them.

(I) is sufficiently supported by some of the examples seen already in §4.1. Lewis's example of the railroad won't do, since this scenario involved multiple beliefs, and Lewis claimed that as soon as all his beliefs were called to mind at the same time ("[o]nce the fragmentation was healed") 'straightway' the contradiction disappeared. But Lewis's mathematical example of the construction with ruler and compass, or Field's of the non-existence of a continuous function from the unit interval to the unit square, seem to present cases where one

[14] Though my arguments leave open that for some modality M there may be 'elementary' instances of M-impossibility which are always impossible to entertain. I remain neutral on this issue in the book.

can coherently form a single belief in a contradictory conjunction without any worries that the belief must evaporate.

For example, one can imagine a very sophisticated mathematician reading the conjunction of relevant set theoretic axioms (and relevant definitions) with the statement about the non-existence of the relevant mapping, understanding the statement as a whole, and assenting. It seems that such a sophisticated thinker can 'hold the whole content' of the conjunction in their mind, and affirm it. Note that since 'logical impossibility' would entail impossibility relative to all other alethic modalities, any one logical example of this kind suffices to establish the general claim about alethic modalities made in (I).

Now, one could try to maintain, as I suspect that some of the authors I discussed in §4.1 might, that this is mere illusion. Perhaps the mathematician can't really bring all elements of the conjunction together 'in one consciousness.' Perhaps she merely has the illusion of thinking it, or there are grades of judgment and the most 'proper' form of such judgment is one she cannot achieve with respect to the conjunction. I regard these as avenues perhaps worth exploring in light of other, special theoretical commitments. But I think it is undeniable that these maneuvers should be viewed as theoretically costly. It seems clear that a sophisticated mathematician would be capable of entertaining true conjunctions of comparably high complexity in other circumstances. What would preclude her from entertaining the impossible one? Surely not the fact that it involves multiple conjunctions. As Priest effectively stressed, that is precisely the kind of content that she sometimes must entertain to do her job. And there is every inclination to say, as the mathematician herself surely would upon further discovery, that she initially believed falsely, and not that she had no beliefs at all. Indeed, if Field is right (as I think he is), then rationality can sometimes *impel* the mathematician to the epistemic state she is in. Surely the starting position should be that rationality impels her to have a particular false but reasonable belief—not a pseudo-belief.

It is worth pausing to note an aspect of each of Lewis's, Field's, and Priest's mathematical examples that will be important soon: these examples involve a great deal of *complexity*. The complexity that is at issue is not, or not necessarily, complexity in the structure of the content or its mode of expression (e.g., the fact that the claims would involve several conjunctions). The complexity at issue is in the circumstances described—in this instance, in the relationship between the set-theoretic, geometric, or numerical entities spoken of. It is the

kind of complexity that makes it completely reasonable for someone to entertain, and 'fully understand' the claim without recognizing it is contradictory.

That complexity contributes to our ability to entertain these impossible contents is actually hinted at in Lewis's discussion of his railroad example. This is a case in which three claims are very easily seen to be in tension as soon as they are all entertained at once. And Lewis seems to imply, not implausibly, that the contradiction in his beliefs is maintained in large part, if not entirely, because the beliefs are kept separate. When he considers forming them into a conjunction he says that 'straightway' the contradiction dissipates. What did Lewis mean by this? Was it possible for Lewis to have formed this single contradictory belief? The framework of fragmentation he proposed, when applied within the confines of a possible worlds conception of content, did not allow this, as Lewis himself must have recognized. This is something we will be exploring in greater detail soon. For now it suffices to note that this question is already a little less clear than that regarding the complex mathematical cases. In those latter cases, the contradictory attitudes seem straightforwardly possible, and sometimes even rational, to hold.

So much for (I). What about (II)?

(II) For some but not all alethic modalities M, there are M-impossible contents such that:

- there is a characteristic strong resistance to the formation of single attitudes borne to those contents; and
- the resistance in question tends to *increase* as the complexity those contents *decreases*.

Let me next take the easy part of (II): the "not all" part of "For some but not all alethic modalities …". Nomological modalities will suffice to show this. There is no special resistance to believing (or supposing, or imagining) that familiar things obey laws different than the actual physical laws. One can believe, suppose, or imagine that electrons attract each other, and repel protons. One can believe that gravity is slightly stronger than it in fact is. One can believe that gravity doesn't exist. I think this is more or less straightforward, provided one regards these as metaphysical possibilities. But if help is needed, I think the possibility of such attitudes seems clearest when we consider people ignorant (even grossly ignorant) of the physical laws, and ask what it is possible for them

to believe (or suppose, or imagine) of the laws. The answer seems to be: virtually anything. It will be worth noting that these counter-nomic circumstances are possible to entertain even tough they can be very simple, and 'surveyable'—they needn't involve a high degree of complexity of the kind found in the impossible mathematical cases recently discussed. For example, there is very little such complexity in an imagined case where two uncharged, qualitatively identical particles, alone in close proximity in an isolated and otherwise empty tract of space, attract each other by gravitational force slightly more strongly than particles with their mass in fact do.

As before, we can try to explain any such appearances away. We can claim that these apparent attitudes would not be about the objects of our surroundings, or 'our' elementary particles, or the forces that play a role in our physical laws. Again, I want to allow this as an avenue we can explore, but will maintain that pursuing it should be viewed as coming with theoretical costs. After all, the objects that we know to be governed by laws have a behavior modeled using mathematical constants whose values we only know to some approximation. Surely we are able to form hypotheses about these objects under various alternatives for the value of those constants within the margin of error allowed by our current approximations. The proposed alternative here is in danger of preventing us from doing even that.[15]

Now for the harder part of (II): the "For some …alethic modalities …". Let me note two important things about this claim. First, the resistance it posits among certain impossibilities is said to be *characteristic* of them. By this I mean that the resistance is typically encountered by a normal thinker. It may be that certain atypical thinkers encounter *no* resistance entertaining these contents. Second, the resistance posited is not presumed to be indefeasible. What this means is that even if a typical thinker encounters the resistance in question, we are leaving open that that it may be overcome, so the thinker is able to entertain the content after all.

Principle (II) is thus a serious weakening of the idea that we came across

[15] As I hint at in the previous paragraph, the entertainability of these thoughts seems clear provided we allow that the physical impossibilities in question are metaphysically possible. If this is denied, the matter becomes murkier. Fortunately nothing in the puzzle I want to raise, nor in my explanation of it, ultimately rests on the assumption that physical impossibilities are sometimes metaphysically possible. In particular, the duality will only require that *if* they are sometimes metaphysically possible, we can bear single attitudes towards them in the way I am suggesting.

in §4.1 that what is possible bounds possible thought. (II) posits resistance to entertaining impossibilities only for some modalities, and even relative to those modalities only for some contents. Not only that, but the resistance is posited only for some individuals, and it is not claimed to be insurmountable.

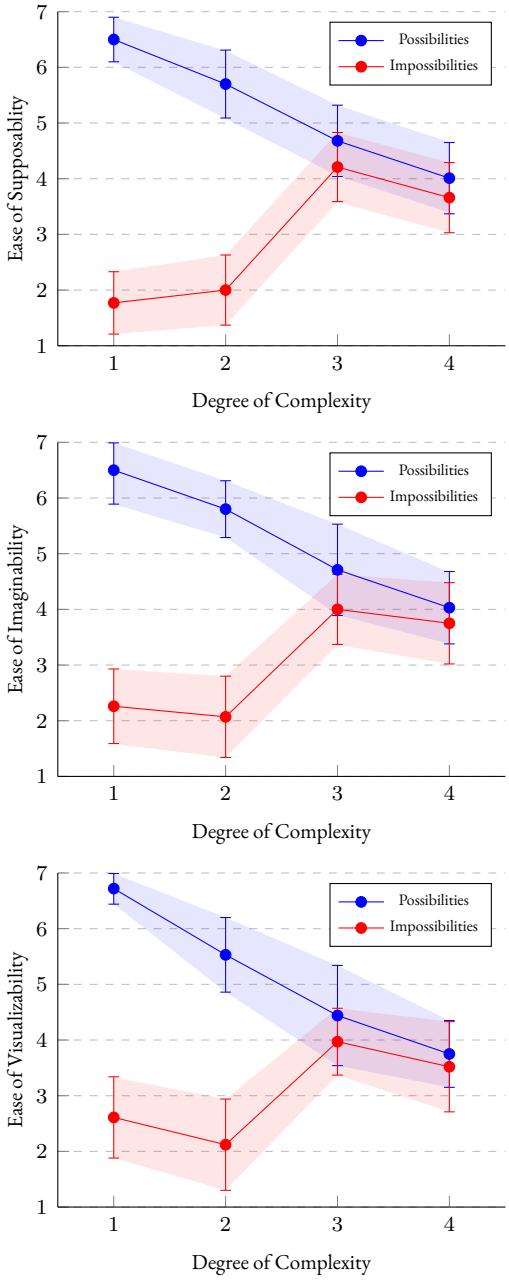
The goal of the weakening is to strike a balance between providing a principle weak enough to be easily defensible, while also being strong enough to require special explanation.

Let me begin by exploiting the weaknesses of my hedging: qualified in the ways I have, (II) is extremely hard to deny. The case for (II) begins by considering resistance to entertaining simple contradictions. The philosophers I discussed often focus on such simple, obvious impossibilities when building their case for the impossibility of illogical thought, typically taking it to be clear that such impossibilities resist entertaining. Whether or not this is true in general, it seems undeniable that it is ordinarily true of the 'typical' thinker.

This is borne out in some experimental work. Figure 4.1 below shows the results of trials where North American consultants recruited through Amazon Mechanical Turk were asked to suppose (N=77), imagine (N=63), or visualize (N=58) certain contents and rate on a Likert scale of 1–7 how easy or difficult they found that task, with 1 being "I could not suppose [imagine/visualize] the scenario" and 7 being "It was very easy to suppose [imagine/visualize] the scenario." Consultants were presented with a randomized set of sentences to suppose (etc.) that were of various levels of complexity, some of which were logical contradictions, others of which were simple possibilities. Complexity here was determined only on intuitively syntactic/logical grounds: a more 'complex' content was expressed with, for example, more conjuncts or the presence of quantifiers.[16] (See Appendix A for more details of the experimental set up.) Figure 4.1 plots the mean values assigned to logical possibilities and logical impossibilities of varying degrees of complexity, alongside 95% confidence intervals (i.e. $\alpha = 0.05$) calculated in standard fashion for an unknown population standard deviation. A complexity of '1' corresponds to the simplest contents—e.g., a simple contradiction. '4' corresponds to the most complex contents (a conjunction of several claims, several of which were quantified).

[16] As per my discussion above, syntactic complexity is *not* the notion of complexity that fundamentally matters to me. But it is a useful quantifiable surrogate, harmlessly appealed to in the experimental context.

FIGURE 4.1: Ease of supposability, imaginability, and visualizability plotted against complexity of content



The blue lines in Figure 4.1 plotting the judgments concerning logical possibilities shows roughly what one would expect. The simplest possibilities are reported as very easy to suppose (imagine, etc.)—close to the extremal value of 7. And we see in the downward slope that an increase in complexity leads steadily to an increase in the difficulty of entertaining logical possibilities. By contrast, logical impossibilities plotted in red seem to behave quite differently. The simplest logical impossibilities were reported as the hardest of all cases to entertain (with the mean closer to the extremal value 1). But as complexity increases, we eventually see a large jump corresponding to reports that these impossibilities have become *easier* to entertain. At this point, the significant differences between possibilities and impossibilities seem by and large to evaporate.

For a simple contradiction (i.e. a logical impossibility of complexity '1'), the median and mode value for the supposability judgments was 1 (this was shared for imaginability and visualizability tests). Indeed, this was the response of a strong majority of participants (roughly 80%). This majority reported that they found it simply impossible to suppose the content they were presented with. The reverse pattern is found for the simple possibility, with the majority (roughly 90%) of respondents reporting a '6' or a '7'.

I take this to substantiate one half of the claim in (II): that "For some but not all alethic modalities M, there are contents p impossible relative to M such that there is a *characteristic* strong resistance against the formation of single attitudes borne to p …". Typically ordinary consultants report simple contradictions are much harder to entertain that simple possibilities—so hard that they often cannot entertain them at all. I have heard some philosophers express a concern that cognitive resistance might only arise for *imagination* or *visualization*, and not for more abstract attitudes like supposition. But North American consultants, at least, seem to exhibit essentially the same form and strength of resistance regardless of whether they are asked to suppose, imagine, or visualize the content in question.

The rest of the data substantiates the remainder of (II): that "the resistance in question tends to *increase* as the complexity of *p decreases*." For the very simplest impossibilities, the resistance is so strong that a strong majority of consultants cannot suppose (or imagine, or visualize) the content at all. Increasing the complexity of contents at some point provisionally *improves* supposability (imaginability, visualizability). Indeed, it improves it quite substantially, until

the standard degrading effect of complexity seems to take back over, and we see only slight differences reported between the difficulty of entertaining a contradiction and that of entertaining a corresponding possibility. Both are reported as more challenging to entertain, but not impossible.

These results raise two critical questions that will occupy me in §4.3: *Why do we find the reported divergence in entertainability between contradictions and possibilities? And why is it influenced by matters of complexity*? These are important question to ask, even if (as both (II) and the data allow) the resistance we find may be defeasible, and may be influenced by matters of expertise or culture (to take just two examples).

Before discussing what a good explanation of this phenomenon should look like, I want to make some remarks about its extent. With the above results, we see that logical contradictions create some characteristic resistance to entertainability. Do any other alethic modalities generate this effect?

The answer appears to be that any modality at least as 'permissive' as metaphysical modality tends to yield this consequence. This is reflected in the historical tradition, where the status I've accorded to strict logical contradictions is often extended to simple metaphysical or 'conceptual' impossibilities such as the fact of 2 and 1 being 4, or the existence of a square circle, a colorless red object, or a vixen that is no fox. On essentially the same intuitive grounds as for contradictions, the resistance to entertainability appears to hold for an importantly broad class of metaphysical impossibilities.

There are two (putative) classes of metaphysical impossibilities that merit special commentary: the a posteriori metaphysical impossibilities discussed by Kripke (1980); and the (alleged) impossibility of a mind existing without any physical or other non-mental basis for it. Let me comment briefly on the former cases here. The latter I will have to set aside here for reasons of space.

The general, defeasible resistance to entertainability seems to exist *even* for Kripkean metaphysical impossibilities, *provided* that a thinker is suitably aware of the contingent facts upon which the impossibility of the content in question depends. Kripke himself often makes this point, repeatedly invoking imaginative resistance in his defense of various instances of essentialism. Here are representative examples from Kripke's discussions of the essentiality of origins and kind-constitution (italicized emphases in the original, my emphases in underline):

> How could a person originating from different parents, from a

> totally different sperm and egg, be *this very woman*? One can imagine, *given* the woman, that various things in her life could have changed … what is harder to imagine is her being born of different parents.
>
> (Kripke, 1980, 113)

> Any world in which we imagine a substance which does not have [the essential properties of gold] is a world in which we imagine a substance which is not gold, provided these properties form the basis of what the substance is.
>
> (Kripke, 1980, 125)

> But whatever we imagine counterfactually having happened to [this object] other than what actually did, the one thing we cannot imagine happening to this thing is that it, given that it is composed of molecules, should still have existed and not have been composed of molecules … [O]nce we know that this [object] is a thing composed of molecules—that this is the very nature of the substance of which it is made—we can't then … imagine that this thing might have failed to have been composed of molecules.
>
> (Kripke, 1980, 127)

> … it is hard to imagine me coming from a sperm and egg different from my actual origins …
>
> (Kripke, 1980, 155)

It is fascinating that Kripke so often appealed to imaginative resistance in discussing cases he took to be metaphysically impossible, especially in light of his acknowledgment of an epistemic modality on which these metaphysical possibilities are epistemically possible. Indeed, on the surface Kripke seems to be motivating metaphysical impossibility *on the basis* of imaginative resistance. Note also that in the third quote above Kripke does not say it is impossible to imagine a given object is made of something other than molecules even if it is actually made of them. Rather he says that *once we know* the object is made of molecules we can no longer imagine it otherwise. So the imaginative resistance

is only claimed to surface, as I stressed above, once one knows the empirical facts underlying the metaphysical impossibility that is relevant.

There are interesting questions about whether Kripke's appeals to imaginative resistance should be taken at face value. But I won't pursue that exegetical question here. For now, I will merely note that it is completely plausible for Kripke to have made such appeals. Once we know that Hesperus is Phosphorus, it is possible for us to imagine scenarios in which we have the same qualitative evidence, and a planet we name "Hesperus" is not the planet we name "Phosphorus". But as Kripke carefully pointed out, this is not to imagine that Hesperus is not Phosphorus. Once someone acknowledges this point, and the fact that Hesperus is Phosphorus, it becomes much harder for them to know what they could do to imagine (or suppose, or visualize) that those planets were not the same.

What this means is that, perhaps bracketing certain exceptional cases like those running up against the mind-body problem, the resistance to entertainability seems to extend with suitable caveats to all metaphysical impossibilities and no further. This will matter, because a good explanation of the cognitive resistance should also include an *explanation of its extent*.

Before starting to look for any such explanation, though, I want to make three short remarks about the scope of the discussed cognitive resistance, and especially the idea that this resistance extends to all attitudes.

First, claims (I) and (II) are only about propositional attitudes, not other cognitive states or activities. I mentioned above a concern I have encountered among some philosophers that the phenomenon of cognitive resistance belongs specially to imagination or visualization, and not necessarily to (say) supposition. While I do think that there are important differences between supposition, imagination, and visualization as propositional attitudes, I also suspect that one source for the concern that we can easily suppose a simple contradiction is a conflation of the propositional attitude of supposition with the activity of 'supposing' something in a form of symbolic manipulation—say, for *reductio* in the context of a particular deduction system.[17] (I suspect Priest may be exploiting this conflation in his quotation from §4.1.) This latter activity is a cognitive one that may sometimes, or even often, be accompanied by genuine propositional attitudes of supposition. But the activities of sym-

[17] Cf. the distinction between symbolic reasoning and true inference drawn in response to MacFarlane in §3.3.

bolic manipulation, and the 'suppositions' that they sometimes involve, are not of themselves attitude states. I do not deny that one can very easily 'suppose' even simple contradictions in such processes of symbolic manipulation. But the frequency with which 'supposition' is used by philosophers for such symbolic processes can distort our thinking about the mental state of supposition. This makes the judgments regarding supposition held by ordinary consultants, who have less contact with such forms of symbolic manipulation, all the more significant.

Second, though I have focused mainly on imagining, supposing, and conceiving, I hope it is clear that the resistance we find for these attitudes extends to all others. This includes other attitudes of acceptance like belief. It should be obvious that belief resists taking simple impossibilities as objects. One may have misattributed this *merely* to the fact that simple impossibilities are typically obvious falsehoods, and thinkers do not (generally) believe at will, but rather believe what they take their evidence to support. But acts of imagination or supposition are characteristically held at will, and one can easily imagine or suppose known falsehoods. All this strongly suggests that the resistance manifesting for belief has deeper origins. The cognitive resistance extends even to preferential attitudes like desire or hope. In these cases it is especially important to maintain the methodological restriction of considering a single attitude borne to a single content. Ordinary, seemingly rational individuals can have contradictory preferences. You may want to take the new job in one sense (focusing on the increase in salary), but not want it in another (after all, it means uprooting your family). What is less clear is that you can want (in any sense): to both take and not take the job. That is a distinctively confusing kind of preference that goes well beyond what any ordinary case of ambivalence could show.[18]

Finally, I want to stress that claim (II) concerns a resistance in a single atti-

[18] I said earlier that I think the phenomenon we are uncovering for single attitudes extends in some measure to multiple inconsistent attitudes. As my remarks here reveal, I *do not* want to presume it extends in this way to multiple inconsistent *preferential* attitudes. Roughly, I am concerned this fails precisely because these attitudes are not attitudes of acceptance, and so bear the structure of something like a ranking over propositions. In Chapter 5, I will make substantial use of the assumption that cognitive resistance extends to multiple inconsistent attitudes in defending a reduction of deductive inference. It may be worth flagging ahead of time that the exemption I am carving out for preferential attitudes here has no bearing on the tenability of that reduction: as noted in Chapter 2, inference does not mediate between preferential attitudes.

tude being borne 'directly' to an impossible content p. This is not yet to make any claims about attitudes borne to contents related to p, or even those taking p as a 'constituent' (if contents have constituents). For example, (II) does not make any commitments about whether the negation of p exhibits any resistance to thought. Nor does it take a stand on whether the claim *that someone believes (or supposes, or imagines)* p itself exhibits any resistance to thought. The status of these more complex contents 'subsuming' p is certainly of interest. But we needn't delve into their status to get clearer on the resistance uncovered for taking a single impossible proposition as an attitudinal object.

4.3 Representational Crowding-Out

The reflections of the foregoing section and the consultant judgments exhibited in Figure 4.1 reveal a kind of duality in the space of thought. For most contents, the simpler the content the easier it is to entertain. But for a special subset of contents—the metaphysically impossible ones—it appears the reverse holds: the simpler the content, the *harder* it is to entertain. This gives rise to a pair of interrelated questions.

- Why do we find any divergence in reported resistance to the entertainability of metaphysical impossibilities and metaphysical possibilities?
- And why does a decrease in complexity of metaphysical impossibilities seem to result in an increase in the resistance to entertaining them?

My goal in this section is to outline the *form* that an answer to these questions should take, and sketch some tools and analogies that I think will be helpful in that task. As will become clearer soon, I think a full answer to these questions requires settling claims about the grounds of mental representation that are far too complex and controversial to tackle in the scope of this book. What is reassuring is that we can place some important constraints on what answers to these questions will look like. And it turns out these constraints will provide lessons enough for the applications I want to explore in Chapter 5.

Let me begin with the second question above. Why are consultants reporting that any resistance in entertaining thoughts become stronger the simpler the thoughts become, and weaker the more complex they become?

One option is to explore an error theory about consultant judgments of complex impossible contents. On this view, complexity has no influence on

whether a proposition's impossibility generates resistance to entertaining it. That resistance is always absolute and independent of complexity. Instead, on the error theory, the effect that increases in complexity have is to make it increasingly likely that consultants fail to follow the instructions of the task set before them. That is, although consultants report that it is easier to entertain complex metaphysical impossibilities, perhaps this is simply because they are entertaining, or trying to entertain, a content other than that they have been instructed to. If the content is suitably complex, it might be easy to confuse the task of entertaining that content with a similar, metaphysically possible content. This error theory could also help explain why there is relatively little difference between impossibilities and possibilities at high degrees of complexity.

While I don't doubt that a failure to recognize an impossibility *as* an impossibility is importantly connected with the judgments we see for high degrees of complexity, I find this brand of error theory implausible for two reasons.

First, the error theory actually requires *two* kinds of mistakes to be made by consultants, and to be made together systematically. That is, we have to suppose not only that consultants are mistaken about what propositions they entertain for high degrees of complexity, but that they are *mistaken in thinking that they succeed* (with difficulty) in doing what is asked of them. Consultants who encounter a content so complex that they *recognize* they are having trouble keeping it in mind have a natural option to choose: the value 1. For if a content is so complex that one can't keep track of what one has been asked to suppose, that means one hasn't yet succeeded in supposing it (even with some difficulty). If complexity makes it so hard to keep track of what one is supposing that the grand majority of consultants are failing to do so, why are so many of them unaware that they are failing (or aware, but not reporting that they are aware)? Why isn't it transparent to them that the content is so complex that they are losing a grip on the instruction so that they should advert to much lower values in reporting how hard it is to suppose what is asked of them?

The second reason to be suspicious of the error theory is because of a point made by Lewis and Priest above: it seems easily possible for a reflective and competent thinker to entertain complex metaphysical impossibilities that are even recognized by them as such. Lewis claimed that he could imagine the construction of a regular polygon of nineteen sides with ruler and compass, even though he knew this to be impossible—indeed, even though he was in posses-

sion of a proof of its impossibility that he no doubt understood perfectly well. Priest claimed he would have no trouble conceiving that Goldbach's Conjecture was true were it proved false, or false were it proved true. It is, I think, implausible to claim that Lewis or Priest would simply be imagining or conceiving of the wrong things. Rather, it seems much more plausible that matters stand as Lewis suggests: to imagine the possibility he does not imagine "arc by arc and line by line." Priest wouldn't imagine the details of a proof.

Once it is acknowledged that one can count as supposing, imagining, or conceiving of highly complex scenarios without entertaining details in this way, there is no special reason to discount the verdicts of consultants. Indeed, because we have good reason to think that complex metaphysically impossible cases can be entertained reflectively and with relative ease, there is every reason to think that it is precisely this 'ignoring the details' that is facilitating ordinary judgments for complex metaphysical impossibilities.

For both of the foregoing reasons, I think we should take consultants' judgments about complex contents, including complex impossibilities, seriously. Incidentally, I should add that we have even stronger reasons to take the judgements for the simple impossibilities at face value as well—precisely because they are simple. There is no hope in saying that ordinary consultants are 'entertaining the wrong contents' when asked to suppose very simple impossibilities. (What contents could they accidentally be trying to entertain instead?)

Once an error theory of the judgments concerning impossibilities is ruled out, we need some explanation of why complexity really does influence how hard it is for consultants to entertain the very contents we ask them to. Since the resistance potentially varies from agent to agent, and appears to constrain their cognition, the most natural explanation of the data seems to be that the agents in question are in a *cognitive state* which explains their judgments. This is the hypothesis I will explore in the remainder of the section.

What exactly is this cognitive state? Well, to understand what the state is, we should understand what it does. Based on the above data and our ancillary philosophical reflections, we should say that for at least some metaphysically impossible propositions p, there is a corresponding cognitive state with the following properties.

(1) *Being in the state impedes one's ability to entertain p.*

(2) (a) *The presence of the state is sensitive to the complexity of p.*

The more complex the impossibility, the less likely one is to stand in the cognitive relation to it; the less complex the impossibility, the more likely one is to stand in a relation to it. This is required to account for the boost complexity gives to the ease of entertaining impossibilities.

(b) *The state is cognitively demanding in this sense: only for the* very simplest *impossibilities is the relation pervasive among typical thinkers.*

What we seem to find is that for extremely simple contents (conjunctions as with complexity 1, or disjunctions of conjunctions as in complexity 2) where, intuitively, a present contradiction is 'easily seen or noticed,' we get significant numbers of ordinary consultants reporting they are unable to entertain impossible contents. As soon as the complexity makes it challenging or unclear where a contradiction resides—where it takes reflective work to 'see it'—we get a substantial drop-off in the number of consultants who seem to encounter it.

(3) *The state is not an ordinary propositional attitude. In particular, it is not the state of belief or even knowledge that p is impossible.*

We can see this from both Lewis's and Priest's examples. Lewis had both the belief and even the knowledge that his ruler and compass construction was metaphysically impossible. This did nothing to hinder his entertaining its existence imaginatively. And as Priest stressed, learning Goldbach's conjecture was false would not prevent him from imagining or conceiving of its truth. If we take these claims seriously, as I think we should, then whatever the source of resistance to entertaining impossibilities is, it cannot be belief or knowledge. Belief and knowledge that a proposition is impossible can persist while the barrier to entertaining the proposition is removed through the influence of complexity. What is more, knowledge and belief are the only plausible propositional attitudes to play the role of the relational cognitive state we are looking for. Other attitudes of acceptance seem to be significantly worse candidates. It follows that the state we are looking for is no propositional attitude.

(4) *The resistance to entertaining p created by the state tends to be 'absolute' when present.*

> We see that when the resistance is encountered by consultants, they tend report it as hindering the ability to entertain a proposition absolutely. That is, consultants choose a value corresponding to an *inability* to entertain a content. This is a level significantly below the levels of difficulty they report in entertaining even extremely logically complex contents. Where the resistance ceases to be encountered at higher levels of complexity, there appears to be relatively little difference between impossibilities and possibilities. And if we follow the verdicts of Lewis and Priest, this is to be expected: once complexity allows for the entertaining of an impossible content, it is not clear the impossibility plays any special role in obstructing one's ability to entertain the content. This seems to suggest that, at least on the whole, the kind of resistance to entertaining that impossibility generates is an 'all or nothing' matter.

When an agent is in the cognitive state that satisfies (1)–(4) in relation to some proposition p, I will say that the agent *representationally crowds-out* p.

Of course, to say there *is* some such cognitive relation is not to get very far in explaining it. In the remainder of the section I would like to make a proposal for how to understand crowding-out states. A core idea behind the proposal is that crowding-out states are not separate cognitive relations over and above a given representation in (say) a propositional attitude. Rather it is a *mode* of a pre-existing representational state—it is the *way* in which that state comes to represent. Such representational modes correspond to rules which pair states of a representing system with represented states. These rules or pairings are not further things represented, but rather the means by which any given act of representation takes place.

To explain this relatively abstract idea, it will be helpful to see how crowding-out relations arise organically in the context of visual imaginative representation to the extent that imagination involves something like pictorial or imagistic representation. Reflecting on features of pictorial representation can not only explain why we would sometimes encounter a resistance to imagining certain impossible propositions, but also why the resistance is not the product of a further attitude beyond imagining itself, why the resistance is sensitive to matters of complexity, and even why the resistance tends to track metaphysical impossibility in particular. Above all, what is most important is that the explanation proceeds 'from the ground up'—by first getting clearer on the conditions on imaginative representation, and then seeing how those con-

ditions are thwarted precisely through the combination of a proposition's being metaphysical impossible and a thinker imagining in a particular way. While the explanation I give for imagination can't extend unmodified to other attitudes (like supposition, belief, or desire), it gives some indications for how an extension to those attitudes could be developed.

So let's ask: how do we represent in imagination? On a natural and familiar picture, one represents through imagination, constitutively, through a kind of 'imaging.'[19] Imagination has a sensory, or quasi-sensory component. One imagines visually or auditorily, for example. And there seems to be a relatively clear role that this sensory or quasi-sensory aspect of imagination plays in explaining its representational properties: it is a prerequisite to imagining some scenario that elements in the sensory representation correspond to something like 'how things would seem' (or could seem) were the scenario one imagines to obtain. A correspondence of some kind is required between the elements of the 'image' of imagination, and what is 'imaged,' for the representation to take place. Here, as I alluded to before, I'll focus on visual imagination, though nothing in the account I give relies on peculiarities of the visual sense modality.

Now as Lewis noted, we need not visualize every detail of an imagined scenario to count as visually imagining it. So more is needed to explain what one visually imagines than a simple account of the elements of one's visualization. Visualized elements can be imprecise and incomplete, yet sometimes count as part of a representation of a content associated with only one of several precisifications or completions of the visualization. What added conditions do the determining? I won't need an answer to this question here.[20] It suffices for my purposes to note that it seems like a necessary, but generally insufficient, condition on imagining a situation *that p*, that there be *some* sensory or quasi-sensory elements 'pictured' in one's mental act that correspond to how things would look (or how things might look) were p to obtain. From this condition alone, we can see why there could be a *mode* of imagining which would give rise to the features of representational crowding-out. This mode of imagining would be a kind of clarity, precision, and completeness underlying the imaging component of an imaginative act.

Consider someone who is imagining a plane figure (like a circle or triangle)

[19] Perhaps not all imagination proceeds in the way I'm about to describe. It suffices for my purposes if some central and important class of imaginative acts do.

[20] For an exploration of some possibilities, see Peacocke (1985), Kind (2001), Noordhof (2002), and Kung (2010).

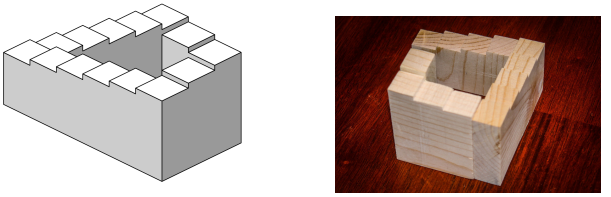
by imagining all its parts together in a single visual image. And they do this with a 'customary' understanding of the image's mode of projection. Were they to imagine a figure they would do it, as Lewis says, "arc by arc and line by line." This gives us the representational mode adopted by this individual in visual imagination. Could a person imagine a square circle while maintaining this representational mode? No. But why? Well, for this person to imagine a square circle while respecting the stated representational mode would require there to *be* a square circle (in their visual image), and perhaps a *possible* square circle (that would be the result of projecting their image). Since neither of these things is possible, neither is representing the square circle in imagination *as this agent is representing*.

The basic point can be understood here without considering mental representation, and focusing instead on pictorial representation. In the purely pictorial case, a representational mode—a method of pairing of representing states with represented states—essentially consists in a projection scheme: a rule which maps points of a picture onto a scenario in the world pictured.

Can there be a picture or image—in a book, say—of a square circle? More generally can there be such an image of an impossible figure? While there can at least *appear* to be images of impossible figures or objects, the only way in which these images can count as representing impossibilities is by adopting 'non-realistic' modes of projection (e.g., by cobbling together incompatible, but independently realistic modes of projection for different parts of the image). For example, a drawing of Penrose Stairs (Figure 4.2 (left)) seems to depict the impossible situation of stairs which descend indefinitely within a finite space. Perhaps it can be said to sometimes actually represent this. But of course, there are consistent, realistic modes of projection on which the two-dimensional image could equally be a representation of a perfectly ordinary (and so metaphysically possible) three dimensional object, like a 3d model of Penrose stairs (Figure 4.2 (right)) used to replicate the drawing's confusing visual effects via forced perspective. As I say, the object depicted in this way is perfectly possible (we have made such objects, and photographed them). This shows that corresponding to the image there is at least one 'realistic' mode of presentation according to which the image (and images like it) would correspond to metaphysical possibilities, and only those.

Even more simply (to consider an example we will return to again shortly), we could take the image on the left to be a representation of a *another two-*

FIGURE 4.2: Penrose Stairs



dimensional image (for example, in another book) with a standard mode of projection. Thinking of the image in these terms makes clear that not only this image, but any such image, would represent a metaphysical possibility (indeed: one effectively witnessed by the image itself).

It is only by taking the modes of projection to be different than these 'realistic' ones that the image on the left of Figure 4.2 could be one of an impossibility. With a consistent, realistic mode of projection, not only can *this* image not be a representation of an impossible object, but *there could be no* such image. This is because the realistic mode of projection, coupled with the image, would give instructions for creating a metaphysically possible object. The metaphysical possibility of the object would be secured, as it were, by the image and its projection.[21] The problem here is just a generalization of the problem with a square circle. You can't make a drawing that represents a square circle using an ordinary, uniform understanding of how the drawing represents, precisely because if you could draw it with that understanding, a square circle could exist (as the projection of your drawing). Since a square circle cannot exist, you can't represent one in the relevant way either.

To return to imagination, the basic idea behind the proposed account of crowing-out relations as they arise in imaginative acts is that the conditions on imaginative representation subsume something like those on ordinary nonmental pictorial representation, at least to some extent. And to that extent, they inherit obstacles to the representation of impossibilities and for essentially the same reasons.

Now, consider again our imaginer. *If* they imagine using the representational mode I described, *then* they will be unable to imagine a square circle. But

[21] Cf. WITTGENSTEIN (1922, §2.203): "A picture contains the possibility of the situation it represents."

are they forced, in any way, to use this representational mode? They may well be. Perhaps they are under a psychological compulsion to represent in this way. Or perhaps they are under a 'conceptual' compulsion to do so: perhaps our account of the nature of imaginative representation will require, for the imagination of *some* simple figures, that to count as imagining them at all one must visualize them in detail, line by line and arc by arc in a 'realistic' mode. But it is important to know that even if someone is driven to this representational mode for some impossible figures, whether psychologically or conceptually, they will not *generally* be driven to that representational mode in imagination. This is, minimally, because of the point made above that imagistic elements generally only do part of the representational work in imagination, and incomplete or indeterminate images used in imagination can represent complete determinate scenarios through the presence of other conditions.

This explains why the resistance to imagining impossibilities is sensitive to the complexity of what is imagined. Perhaps one will not, or cannot, count as imagining a simple lone circle unless one visualizes (say) the entire thing clearly. But this is not true of other figures. There is no psychological or logical compulsion to visualize a figure like a chiliagon (another example of both Lewis and Priest) line by line in order to imagine it. We can 'leave out the details' and still count as imagining that thousand-sided figure. (Though perhaps an imaginer of truly extraordinary, inhuman abilities could be able to visualize the details in such a case. They may even be psychologically compelled to do so.) The conditions on representation in imagination thus allow latitude in degrees of accuracy and detail when what is represented is correspondingly complex. And the more complex, the more latitude: the less work is permitted to be done by the imagistic representational elements of imagination. And the less work done by these elements, the more room may grow for the imaginative representation of an impossible scenario. That would explain how Lewis successfully imagines his ruler and compass construction.

Basic reflection on the conditions of representation in imagination thus *yields* states of crowding-out in the form of certain imagistic representational modes. These are roughly the clear, complete, precise representations with realistic modes of imagistic projection. Critically, these relations are not *further* cognitive states over and above imaginative acts, but rather certain *ways* those imaginative acts take place. Moreover imagistic representational modes have precisely the features we want of states of crowding-out.

For example, treating imagistic representational modes as crowding-out certain propositions helps us understand why there is resistance to imagining impossibilities in imagination at all (as per condition (1) above): sometimes the conditions on representing in imagination, coupled with a given imagistic representational mode, *render the grounds of the impossibility of some circumstance the selfsame grounds as the impossibility of representing it*. In the process, the account on offer shows us why crowding-out relations tend to create an *absolute* (though in-principle surmountable) bar to representation of a proposition (as per (4)). It further helps us understand why *metaphysical* impossibility is what characteristically generates obstacles: if one represents using the representational mode in question, this guarantees that what one represents is possible only in the broadest sense of possibility. (For example, we could imagistically represent things that are physically impossible.) So if something is impossible in that broadest sense—the metaphysical sense—one could not represent it. The account helps us understand why complexity matters (as per (2a)): increased complexity in what one imagines tends to relax the need for, or prevalence of, the representational mode in question. Moreover the account helps us understand why the cognitive relation is cognitively demanding for ordinary humans (as per (2b)): only the most absolutely simple scenarios are ones that ordinary humans *can* image with the level of detail and accuracy demanded of the representational mode. Even modest increases in the complexity of what one imagines will, therefore, ensure the relevant representational mode tends to dissipate. The account also explains why some propositions, but not others, might be blocked in imagination. For example, one can adopt a very precise representational mode for one figure, but not another, in a single imaginative act. And finally the account allows us to see the cognitive relation is not an attitude like belief or knowledge, or even a further cognitive state of any kind (as per (3)). Rather, as I've emphasized, it is in the *way* one represents in imagination that we can locate the sources of imaginative resistance we are looking for.

I propose to think of the case of imagination as a model for giving a satisfying account of crowding-out relations in propositional attitudes more generally. What we want is to give an account of the nature of representation for those propositional attitudes that organically gives rise to representational modes of such attitudes that create the representational resistances we sometimes encounter. Though I favor this explanatory strategy, I will not be able

to pursue it in great detail here. The range of views on how attitude states like supposition or belief represent is vast and remains highly contentious, and there is accordingly no hope of adequately exploring that range of views here.

I am content to leave matters there for two reasons. First, we have some data on hand that any theory of representation needs to account for. The data strongly motivates the existence of some cognitive state satisfying conditions (1)–(4) above. Whatever that state turns out to be, it will be capable of doing most, if not all, of the work that I will set crowding-out relations to in the coming chapters. Some foundational accounts of representation may have trouble accommodating the existence of crowding-out states. But if they do, that is ostensibly a problem for those views of representation, not for the existence of crowding-out states. Second, although there are very important disanalogies between visual imagination and other representational attitudes like belief, supposition, and desire—notably, that the latter needn't have an imagistic component—nevertheless there are good reasons to think that the abstract shape of the account from imagination could be extensible to the other attitudes.

How so? Well, the key idea behind the treatment of imagination is that at *some* level of abstraction, successful representation tends to involve components that *mirror* components of what is represented. In imagination the mirroring is almost literal, and comes in the form of images that somehow match reality or its appearance via a projection scheme. But there are good reasons to think that all representation could share this feature with imagination, even if at a much higher level of abstraction.

In fact, barriers to the representation of impossibility are already familiar from a number of accounts of the grounds of mental intentionality for essentially these reasons. To take one example, on the information-theoretic conception of mental representation (Dretske (1981), Stalnaker (1984)), mental states represent by 'indicating' the truth of their propositional content. That is, an attitude state takes a proposition as object because that states covaries under ideal circumstances with the truth of that proposition. Familiarly on this view, it immediately falls out that it is impossible for an attitude state to take a metaphysically impossible content as object. This is because no state could covary under any circumstances with the truth of such an impossibility, since the impossibility is true in no possible circumstances for the state to covary with.

To illustrate the idea behind this feature of the information-theoretic ac-

count, consider a familiar set of analogies: on the information-theoretic view, mental states represent reality in something like the way that a thermometer indicates ambient temperature or the rings on a tree indicate its age. In normal circumstances, the level of mercury in a thermometer covaries with changes in temperature, and the number of rings on a tree covary with its age in years. Note that the way in which a thermometer or the rings on a tree indicate temperature or age is strikingly different from the way a picture represents a scene. Even so, at a high level of abstraction we can see indication relations as possessing a similar 'mirroring' quality, which is why we find a similar 'resistance' to an indication of the impossible. What would it be for a thermometer to indicate an impossible temperature? Or for the rings on a tree to represent an impossible age of that tree?

Of course, the information-theoretic account unalloyed cannot be the whole of our story of mental representation, not least because it cannot account for the differential influence of complexity on entertainability. This even seems to hold when the information-theoretic view is paired (as it increasingly is) with the tools of mental fragmentation discussed in §4.2. Fragmentation can start to explain the influence of complexity in the form of the cognitive load involved in integrating 'fragments.' But as already discussed earlier, fragmentation only allows for multiple attitudes to be borne to several metaphysically incompatible contents, not a single attitude to be borne to a single metaphysically impossible content. So there is more work to be done in this sphere. Even so, the view still gives us an idea of how a view of mental representation that does not depend on the idiosyncrasies of imagistic representation could nonetheless organically give restrictions on the entertainability of metaphysical impossibilities. And this is one example among several.[22]

So far I've noted that there is a puzzling duality in the space of an agent's thought, posited cognitive states of crowding-out to account for that duality, and sketched an account of those states as representational modes that should arise organically from our accounts of the nature of mental representation. Both in positing states of crowding-out and sketching an account of them, I

[22] My own preference, right now at least, would be to explore the foundational approach to semantics I attribute to Wittgenstein in Shaw (2023) on which an explanation of the grounds of impossible representations would trace to what Wittgenstein calls the "philosophical grammar" of representational talk and its significance. The issues I am discussing in this chapter are, of course, ones that were very much on Wittgenstein's mind as he worked in the foundations of representation both in the *Tractatus* and in the *Philosophical Investigations*.

have mostly tried to stick to what the data from consultant reports motivates, only occasionally drawing on further observations of philosophers like Lewis and Priest.

In what remains I want to go beyond the data in three respects that will be important for subsequent chapters. To begin, I want to defend the following two connected ideas.

(5) It is possible to representationally crowd-out some metaphysically *possible* propositions.

(6) *Given* the purpose of properly mirroring the space of metaphysical modality, states of crowding-out are accuracy-assessable: a state *accurately* crowd-outs p if and only if p is metaphysically impossible.

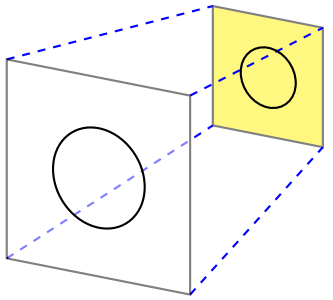
Here, instead of leaning on data from consultants, I will appeal to the particular account of crowding-out I have put on offer, leaning more heavily on analogies I've been developing with pictorial representation. To that extent, the justification for (5) and (6) will be more tenuous. But these claims are also meant to gain further important justification through their applications both in this chapter and the rest of Part I.

Above I defended the claim that complexity really does relax the cognitive resistance created by metaphysical impossibility. If that is right, there are many impossible propositions p such that one may not crowd-out p in a representational act. This means representational capacities can *over*generate in how they implicitly characterize the space of metaphysical modality. The basic idea behind (5) is that a system that can overgenerate in these ways would also be capable of *under*generating as well.

In the case of pictorial representation we can see relatively clearly how both over- and undergeneration of representational possibilities arise. To do so, let's take a highly simplified case in which we are using markings in ink on a white sheet of paper to represent patterns of ink which appear on a second yellow sheet of paper of the same dimensions. (Perhaps the yellow sheet is in another room, and you are curious how it looks. I draw on the white sheet to show you.) I will be assuming the sheets have a continuous spatial structure and the same dimensions. In this context, the most natural mode of pictorial projection—which in this context generates our representational mode—would map each point on the first sheet to its natural corresponding point on the second sheet as in Figure 4.3. (So, corners of the white sheet to corners of

the yellow sheet, the center of the white sheet to the center of the yellow sheet, etc.)

FIGURE 4.3: Pictorial Projection Yielding Accurate Space of Possibilities

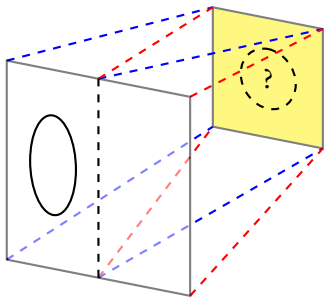


It is easy to see that the 'representational space' afforded by this mode of representation accurately mirrors the possibilities for patterns on the yellow sheet. That is, there is a 1-to-1 correspondence between possible represent*ing* states of the white sheet and metaphysically possible states of the yellow sheet that are represent*ed*. Accordingly, it is metaphysically impossible, while using this projection, to represent a metaphysically impossible state of the yellow sheet, just as I noted above: an 'impossible state' of the yellow sheet could only correspond to an impossible state of the white sheet. In this sense, this representational scheme crowds-out all and only propositions describing an impossible state of the yellow sheet.

But there are alternative modes of projection for which this is not the case. For example, there are modes of projection in which some point on the represented yellow sheet corresponds to *two* or more points on the representing white sheet. A simple projection doing this would be one that separately maps each of the left and right halves of the white sheet to the entirety of the yellow sheet as in Figure 4.4.

If we use this projection, our white sheet can now represent contradictory information about the yellow sheet. Whereas one part of our white sheet may have an ellipse (representing a circle centered within the yellow sheet), the other half may remain blank (representing the *absence* of any figure, including any circle). So our mode of representation now overgenerates: there are more representations of configurations on the yellow sheet than are in fact possi-

FIGURE 4.4: Pictorial Projection Overgenerating Space of Possibilities

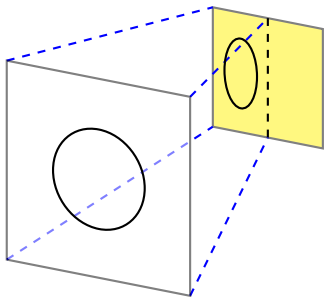


ble.[23] Every drawing on the white sheet on which the left and right halves are not identical attributes impossible combinations of properties to the yellow sheet. This pictorial projection does not, however, undergenerate: for each metaphysically possible distribution of ink on the yellow sheet there is a possible state of the white sheet that represents it—one on which the left and right halves of the white sheet are identical. For example, to represent a circle on the yellow sheet, we would inscribe two identical ellipses on either half of the white sheet. So like our first representational mode, this scheme crowds out no propositions describing possible states of the yellow sheet. Unlike that first scheme, this new one ceases to crowd out some propositions describing impossible states of it.

Even so, we can concoct undergenerating representational schemes as well. For example, we could take a projection which maps the representing white sheet onto only half of the yellow sheet, as in Figure 4.5. In this case, our pictorial representational scheme leaves us helpless to characterize variation on the right side of the represented white sheet. How does a given configuration of ink on the white sheet 'represent' the right half of the yellow sheet as being using this projection? This is in effect a further specification of the representational mode. The particular implementation that interests me here is one on which this half of the yellow sheet is treated as blank by representations on the white sheet. On this scheme, one might say, the projection implicitly characterizes the right half of the yellow sheet as 'necessarily blank' or alternatively as

[23] While this example may seem artificial, this is essentially how two-dimensional 'representations of impossible objects' like the Penrose Stairs operate—by exploiting subtle shifts in our tendencies to extrapolate information about (say) perspective or distance.

FIGURE 4.5: Pictorial Projection Undergenerating Space of Possibilities



'impossible to be filled in.' It does this by not creating any space in the scheme for a representation of anything other than blankness on the right side of the sheet. Note that in spite of its deficiencies, this representational scheme does not overgenerate: every way of filling out the white sheet corresponds to a genuine possible state of the yellow sheet. So this scheme crowds-out all propositions describing impossible states of the yellow sheet, and some propositions describing possible states of the yellow sheet besides.

So various uses of projections in pictorial representation give rise to 'adequate' representational pictorial schemes (Fig. 4.3), but also overgenerating (Fig. 4.4) and undergenerating (Fig. 4.5) schemes as well. The idea I would like to take on as a working hypothesis is that pictorial representation is not unique in these respects, and that mental representational modes may admit of these features as well. We already have evidence that there can be adequate modes of mental representation that crowd out the representation of impossibilities, as well as inadequate modes of representation that proliferate them. This all makes sense if representation at *some* level of abstraction operates as a kind of mirroring of what is represented (where, e.g., information-theoretic accounts of representation give one simple example of how this mirroring relation might hold). The pictorial case gives us a model for understanding how this mirroring can give rise not only to adequacy and overgeneration, but undergeneration as well. As I say, it is a hypothesis that undergeneration may occur in a mental representational scheme. But the hypothesis is natural given what we do know about relations of crowding-out states. Moreover, as we will shortly see, the hypothesis allows for some important and needed flexibility in

our understanding of such states.

I have already surreptitiously introduced talk of 'adequacy' in discussing representational modes and so the crowding-out states they engender, as per point (6). I hope it is relatively clear how and why one might apply the terminology of adequacy or inadequacy when one surveys the schemes like those in Figures 4.3–4.5: intuitively some of these are more adequate representational modes than others. But I also want to be suitably cautious about the interpretation of this talk. Though my language and examples may suggest it, I do not want to commit myself to the idea that representational modes are normatively evaluable *of themselves*.

This may be surprising. Isn't something going *wrong* with schemes that over- and undergenerate? I want to stress that there is something going wrong with the schemes, but *only* given a 'purpose' for them. One such purpose is, of course, to allow no more and no fewer represented possibilities than there are. And by that standard, schemes like those in Figures 4.4 and 4.5 are defective or inadequate. But one may develop a representational scheme with many aims in mind, and adequacy *in the foregoing sense* many not figure at all, let alone decisively, among them.

Consider that the second scheme of Figure 4.4 might naturally be used as part of a mechanism in which images from each half of the sheet are separately seen up close by each of two eyes of a creature like a human with binocular vision. Perhaps the mechanism exploits shifted projections generating stereopsis when a three dimensional scene is represented, but eschews shifting projections when a two-dimensional scene is represented. In this context, something like an overlapping scheme of projection might be the perfect one for the purposes at hand. Or consider that the third scheme of Figure 4.5 could be used in a context where there was already good information about the right side of the yellow sheet—e.g. we are already *certain* it is blank—and we want to use the limited resources afforded by our representing white sheet to magnify details on the left side.

We should also bear in mind that even when we do evaluate a representational mode, it is never evaluated in the same way as a representation itself. For example, even if the overgenerating and undergenerating schemes are part of failed attempts to allow no more and no fewer represented possibilities than there are, nonetheless *representations using these schemes may be perfectly accurate*. It may be that we produce a representation using either of these schemes

that is a perfectly faithful representation of which features the yellow sheet *actually* has (consider the example of representing a circle using two ellipses above). In this context, the representation itself is unexceptionable even if there is something faulty with the scheme. Conversely, and quite obviously, one can have an 'accurate' scheme for the purposes at hand (say in the form of the projection of Figure 4.3) which represents inaccurately (for example, a square is inscribed on the white sheet whereas a circle is inscribed on the yellow sheet).

Whether a representation is accurate or not depends in part on the scheme used, in the sense that the scheme settles what is represented to begin with. But once a scheme is settled we can speak of the representation as correct or incorrect. And whether it is correct or not does not depend on any further aims we had in using the representational scheme. By contrast, the scheme itself cannot be evaluated independently of the purposes to which it is set.

What is more (as will be important for Chapter 5), we can sometimes evaluate acts *based on* a representational scheme fulfilling presupposed purposes. For example, suppose I erroneously use the scheme of Figure 4.5 in an attempt to mirror the space of genuine metaphysical possibilities for the yellow sheet. Then if because of my problematic scheme I *judge* that it is metaphysically impossible for the right side of the yellow sheet to be anything but blank, then the judgment can inherit its flaws from the failure of the scheme to fulfill its purpose.

I submit again that what holds of pictorial representation holds of mental representation as well. Whether our mental representational modes are evaluable will have to depend on their having some role or purpose. This is the reason for the important hedge in (6): "*Given* a purpose of properly mirroring the space of metaphysical modality…". Only relative to some such purpose can we evaluate crowding-out states at all. So far, I should stress, I have not said that representational modes *have* such a purpose—intrinsically, as it were, or in our mental economy. If they did, it would have to be argued. I am skeptical that there is such an argument that applies with any generality.[24]

I've noted that points (5) and (6) lean on analogies with pictorial represen-

[24] I am especially skeptical that there could be such an argument that would apply to all attitudes. Surely a purpose in supposing or imagining could be to suppose or imagine something *interesting* or *fun* (say, for storytelling purposes). And in this context couldn't supposing or imagining the impossible be the best way to achieve this result? Indeed, don't we have many examples of stories (e.g. those of Borges) that seem to be aimed at evoking such attitudes?

tation that go beyond what is directly supported by the consultant data. But I also alluded to the fact that these points can gain added justification from how they supply a needed flexibility in our account of crowding-out states. To expand on this, let me consider a very important question that faces the account of crowding-out that I have put on offer: how can we account for *variation* in which propositions are mentally crowded-out, not only by laypersons like those I have taken as my consultants, but by experts as well?

Ever since Aristotle dismissed Heraclitus as not *really* believing the contradictions he avowed,[25] there has been a temptation among detractors of thought-of-impossibilities to dismiss any apparent counterexamples to their claim, even when the counterexamples come from the reports of highly sophisticated reasoners. While I think there can sometimes be reasonable grounds for these dismissals, there is also an *ad hoc* character to consistently employing this strategy as the number and kind of conflicting reports increase. This would be especially true for an account like mine, which is (among other things) aiming to account for the responses we get from ordinary consultants. After all, roughly a fifth of such consultants reported that they *can* suppose a contradiction, and some of them without much trouble.

Fortunately, the framework I've proposed here is well positioned to account for all manner of variation in judgments about entertainability without simply dismissing them. Let me focus on two special cases that illustrate that range of options: the case of the inattentive or obtuse, and the case of the unorthodox theoretician.

I noted that crowding-out states seem to be cognitively demanding. As we saw, small increases in complexity can drastically improve the ease of entertaining impossibilities. The explanation for why the majority of consultants report themselves unable to entertain simple contradictions was that consultants tend to be forced into clear representational modes that crowd-out these propositions, precisely owing to their simplicity. But it is important to note that I did not say, and there is no need to go so far as to say, that the simplicity of the contradiction was such that it *could not but* be crowded-out. That claim is in no way needed to give any of the explanations I've offered so far, including the explanations of the absolute character of the cognitive resistance impossibilities tend to generate.

Because we have already committed to crowding-out states being cogni-

[25] *Metaphys.* 1005b23–25.

tively demanding, we are free to take some or all consultant judgments about successfully entertaining simple contradictions at face value *if we like*. We may simply say that such consultants are in an epistemically weaker position than the larger majority of consultants who encounter the resistance offered by simple impossibilities—a weaker position on which they fail to crowd-out representation of the impossibility. Perhaps this is because the consultants are simply inattentive or rushed. Perhaps it is that they are cognitively impoverished or obtuse. It doesn't matter. Any of these explanations would be enough to explain how such thinkers can represent the impossibility that they are asked to entertain. It is important for the plausibility of this claim that crowding-out p is *not* the belief that p is impossible, nor is it entailed by having any such belief. The claim that some, or even many, ordinary thinkers can be so inattentive or rushed as to not crowd-out a simple contradiction is much weaker than the claim that these consultants can be so inattentive or rushed as to fail to *believe* that it is impossible. As such, the claims I am making are also not hostage to the prediction that consultants who are able to entertain simple impossibilities must likewise judge that what they entertain is possible. After all, judgements (and even knowledge) that the contents are impossible are still compatible with failing to crowd-out the relevant impossibilities.

One might worry, even given this last caveat: is it plausible that a reasonable, ordinary thinker could not crowd-out a simple contradiction? I suspect such a worry is overlooking the fact that crowding-out states are primarily cognitive states *posited to explain* the judgments of consultants and theorists alike. It is an open question how demanding this cognitive relation is, and it is the judgments of human agents (or some subset of their judgments that we end up taking seriously) that will dictate our answer to that question. What this means is that the framework I am proposing has a great deal of flexibility to accommodate any data from entertainability judgments, provided at least that the totality of that data *roughly* follows the patterns seen in §4.2.

So much for the inattentive and obtuse. But what about the unorthodox theorist? What about a Heraclitus, a Heidegger, or a Priest who reflectively and baldly asserts simple contradictions? Supposing we disagree with them, can we at least acknowledge that they are sincere when they assert things which we take to be simple metaphysical impossibilities? Even if we needn't characterize these thinkers as asserting in bad faith, it does not seem like a great improvement to cast them as simply inattentive or obtuse.

Fortunately there are resources to avoid making such charges. Even though crowding-out states are not attitudes like belief, nothing prevents us from saying the presence of the states could be *responsive* to someone's beliefs, and in particular responsive to their reasoning and justification.[26] The unorthodox theorist often has a sophisticated set of reasons which lead them to make the assertions they do, even if they are assertions of simple impossibilities. It is possible that the reasoning which leads a theorist to endorse the possibility of an impossible proposition p could also make them more likely to adopt representational modes on which p is not crowded-out. Reasoning could, at the very least causally, influence the representational modes a theorist employs, and in particular lead to the adoption of modes which overgenerate in ways analogous to pictorial projections in Figure 4.4. One way reasoning could have this effect is by creating the illusion of more complexity in a circumstance characterized by a metaphysically impossible proposition than is in fact present. The idea that there is a kind of hidden complexity to such cases is one of things that drives many theorists to maintain that simple impossibilities are in fact possible after all.

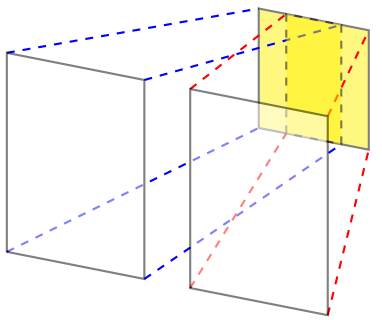
It is worth stressing that not only can this account accommodate the judgments of laypersons and theorists alike who erroneously fail to crowd-out some impossible proposition, but because of the adoption of point (5), it is also able to accommodate cases where agents crowd-out propositions which are not impossible. Putting all this together, we see that we can let the theory adjust to fit the data almost irrespective of what the fine details may be.

I have just one more point before concluding my main discussion of crowding-out states. At the outset I flagged that I would focus on the case of a single attitude borne to a single content. But it is natural to extend the account I've offered of crowding-out states to *collections* of attitudes. The basic idea is that what holds of a single a single representation vis-à-vis the possibilities and impossibilities of representation can hold of multiple representations as well provided they are organized or regulated in a certain way. (We might speak of this as a further mode of representation—a mode of *multiple* representations. Such modes correspond to rules which pair collective states of a representing system with represented states.)

[26] Though, as we will see when we discuss the Adoption Problem in Chapter 5, alongside other forms of rational regress, it will be a much murkier issue whether crowding-out states can ever be *rationally* influenced by reasoning and judgments.

To see this, it again helps to return to the case of pictorial representation. Suppose I represent the state of the yellow sheet now with two separate white sheets. If the portions of yellow sheet represented by each of the white sheets overlap, then we get a *system* of representations that individually do not overgenerate the space of possibilities, but which taken together can overgenerate, as in Figure 4.6. This occurs for essentially the same reasons that we get the possibility of overgeneration through overlapping representational projections schemes for a single representation in Figure 4.4. But just as with

FIGURE 4.6: 'Overlapping' Plurality of Pictorial Projections



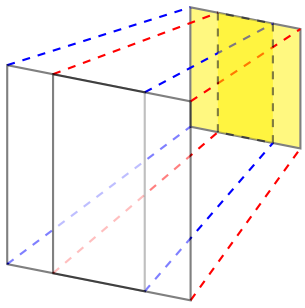
with single representations, overlapping projections for multiple representations can be safeguarded from mischaracterizing modal space if there is 'replication' of the properties of representations for the relevant region of overlap.

What is salient in the case of multiple representations is that there may be ways of trying to 'integrate' the representations to secure the coordination needed to restore a match between the space of representations and possible represented states. For example, suppose that I recognize how each of the two representing white sheets share a region of representation on the yellow sheet. As a result, I 'merge' them so that for this overlapping region any act of altering one sheet results in a commensurate change in the other as in Figure 4.7. (This could work roughly as, e.g., the production of carbon copies).

Now the setting or mode of representation has been adjusted in ways that preclude the representations of impossible states of the yellow sheet. The representations have be 'integrated' so that they function exactly as would a single adequate representation like in Figure 4.3.

Just as we have various models of how pictorial 'mirroring' can occur in the

FIGURE 4.7: Overlapping and 'Integrated' Plurality of Pictorial Projections



mental sphere through views like information-theoretic accounts of representation, so too we already have at least one existing model of how this integration of mental states can occur in the form of techniques of fragmentation that we have already mentioned in discussing Lewis. Roughly, the idea behind fragmentation is just that seen in the case of pictorial representation: fragmented states function like two separate representations, but 'integrated' states function as would a single coherent one. Something in the states, or the way they are regulated, coordinates them so that they function as a representational unity.

The final thesis about crowding-out states that I will endorse is that one can not only crowd-out the representation of a *single* proposition, but that one can crowd out a *joint* representation of several propositions.

(7) A thinker can crowd-out a *collection* of propositions, which interferes with the thinker's ability to simultaneously entertain them, exactly as crowding-out their conjunction would preclude entertaining that conjunction.

Again, claim (7) goes beyond the data that motivates (1)–(4), but flows naturally from the abstract representational framework I've been developing to explain those claims. Here, I will leave off further consideration of that framework, since I think that one of its best defenses comes through the fruits it yields through an extended application in the development of a reductive analysis of deductive inference. It is to this task I'll turn next.

CHAPTER 5

Deductive Inference: Closing Deliberation through Constrained Cognition

In Chapter 2, I gave a skeletal account of inference which yielded a pair of necessary conditions for deductive inferential goodness: a constant modal condition governing the contents in inference, and a psychologically variable condition of 'appreciability.' In exploring the role that inference plays in logical inquiry, I set aside the second condition, emphasizing that its psychological variability explains why logicians would naturally abstract away from it.

In this chapter, I turn back to explore the character of this appreciability requirement. Recent debates, growing in large part from the galvanizing work of Boghossian (2014), have explored how inferential acts seem to bear a specific and unusual set of features, especially in light of the ways that an agent appreciates the goodness of their inferences. It has accordingly been a matter of contention how a mental activity could simultaneously bear those features.

The goal of this chapter is to argue that here we encounter an unusual place where logic can serve to illuminate the nature of deductive inference. Reflection on the logical problem of how impossibility constrains cognition in Chapter 4 led us to posit a cognitive state of crowding-out with various unusual characteristics. It turns out that these characteristics render crowding-out relations ideally suited to flesh out the notion of appreciability required in an account of the nature of deductive inference.

§5.1 surveys the literature on appreciability requirements, using that literature to draw out six constraints on an account of deductive inference and discussing how existing theories struggle to jointly satisfy them. §5.2 briefly surveys received accounts of interrogative states. §5.3 then shows how the combination of crowding-out states and interrogative states can be used to trans-

form the skeletal account of Chapter 2 into a full-fledged reductive analysis of deductive inference, and proceeds to show how this analysis either satisfies, or provides productive routes for satisfying, the six constraints from §5.1.

5.1 Explananda for an Account of Inference

Let me begin by presenting an opinionated overview of some of the recent literature on inference, focusing especially on how we should understand appreciability requirements in good deductive inference. Along the way, I will extract a series of constraints on what an adequate account of inference and appreciability should be able to do. I should stress that all the constraints I formulate in this section will be defended for *deductive* inference specifically. I will return to the question of whether, and how, they extend to ampliative inference in Chapter 6.

Boghossian (2014) reignited debates over the nature of inference by focusing precisely on the unusual constraints required of an appreciability requirement in inference (including ampliative inference). He begins by noting, as we did in Chapter 2, that an inference requires more than merely believing one thing after another. He then reminds that a causal connection between acceptance states is insufficient for inference to have taken place, owing to the presence of 'deviant' causal chains. For example, my belief that there is a tarantula on my arm might have caused the belief that I am scared by first causing fright of which I am aware. This can all happen without my having *inferred* that I am scared from my belief that there is a tarantula on my arm. This gives us a first simple constraint on an adequate account of inference.

(I) *Deviant Chains*: An account of inference should illuminate why mere causation among acceptance states is insufficient for inference to take place, and describe what connection between acceptance states *is* required for it.

Boghossian emphasizes, I think correctly, that progress on this front will require a focus on the rationalizing connections that hold between premises and conclusions in inference. On this basis, he formulates a necessary condition on inference in roughly the same spirit with which I introduced an 'appreciability' requirement in Chapter 2.[1]

[1]Boghossian focuses on conscious, person-level transitions, which certainly informs his framing of the issue of appreciability. Recall that for me appreciability is modalized precisely

> Taking Condition: Inferring necessarily involves the thinker *taking* his premises to support his conclusion and drawing his conclusion *because* of that fact.
>
> Boghossian (2014, 5)

Boghossian does not presume we have some clear, antecedent understanding of the 'taking' involved in Taking Condition. Rather, the thought is that we have an inchoate conception of the roles that this inferential taking must fulfill. So the notion gives us a placeholder that can be fleshed out in various ways by competing accounts of inference. And these accounts can be tested based on how well they answer to the desiderata we are inclined to impose on the relevant 'taking-role.'

In articulating Taking-Condition, Boghossian became a prominent figure in a long and continuing tradition in thinking about inference.[2] Many have embraced the condition, and proposed candidates to fill the taking-role. But even Boghossian's schematic Taking Condition is controversial. For example, Armstrong (1968) and Winters (1983) give accounts of inference in more 'brute' causal or dispositional terms. And Wright (2014) argues against Taking-Condition on grounds of regress, suggesting we instead make progress by seeing inference as an instance of behaving in a certain way for reasons, on analogy with acting for reasons.[3]

In this context, I find the work of Hlobil (2014, 2016b, 2019) helpful and illuminating. Hlobil's work slightly modifies Boghossian's framing of the issue in a way that allows us to skirt some of the controversial elements of Taking Condition while clarifying the challenges faced in giving an account of deductive appreciability.

Hlobil first suggests a methodological reorientation. Rather than focusing on accounting for inferential taking, he notes that conscious inference seems to give rise to a more basic Moorean phenomenon that can be less contentious

because I wish the account to extend to possible sub-personal or sub-conscious inferences. This is one way in which an 'appreciability' requirement as I've formulated it is *weaker* than Boghossian's Taking Condition.

[2] The idea that in conscious inference an inferrer 'takes' their inference to be good finds various forms of expression and endorsement in, for example, Locke (1690/1979), Frege (1879?/1983), Peirce (1905), Russell (1920/1988), Thomson (1965), Deutscher (1969), Stroud (1979), Sainsbury (2002), Field (2009), Dogramaci (2013), Neta (2013), Broome (2014), and Valaris (2014).

[3] For further expressions of skepticism see McHugh & Way (2016, 2018), Siegel (2019).

than Taking Condition, and also cries out for some kind of explanation.[4]

> Inferential Moorean Phenomenon (IMP): It is either impossible or seriously irrational to infer P from Q and to judge, at the same time, that the inference from Q to P is not a good inference.
>
> Hlobil (2014, 420)

Hlobil thinks that as regards *conscious* inferences and judgments this is both straightforward and exceptionless.[5] Note also that the claim holds equally of ampliative inferences and suppositional inferences. It is worth pausing to note just how surprising the latter fact is. It is not clear whether there are *any* general rational constraints on the holding of suppositional states beyond perhaps some loose pragmatic ones (e.g. 'don't waste your time idly supposing things when you should be busy doing the dishes'). But somehow transitioning from one suppositional state to another while possessing a belief about their rational relations *is* subject to a very strong rational assessment. Why?

Even if we construe inference in more causal terms or, like Wright, as an instance of something like acting for reasons, the IMP is still something that we are beholden to explain. It is clear how two beliefs, for example, can stand in rational tension. But how can a belief come into rational tension with an inference, which is no attitude but merely a transition between them? We *may* explain this rational tension by appeal to an 'inferential taking.' But we need not. Whether or not we do, Hlobil's work places a second constraint on adequate accounts of inference.

(II) *Moorean Incoherence*: An account of inference should explain the IMP—that is, how a rational tension generally arises between any conscious inference and the concurrent conscious judgment that the inference is bad.

Not only does the IMP call out for explanation, but that explanation is surprisingly challenging to provide. For example, one popular account of inferential taking is the Intuitional Account, on which inference involves an 'intuition' or

[4] See also the related Moorean phenomenon discussed in Marcus (2012),

[5] For my part, I think it is certainly prevalent and *perhaps* exceptionless. I leave open that there may be very special cases in which irrationality needn't arise. See my discussion of epistemic akrasia in §5.3.

'seeming' to the effect that the premises support the conclusion.[6] But as noted by Hlobil, this proposal faces obstacles in accounting for the IMP. This is because the rational force of 'intellectual seemings' is typically defeasible. Indeed, defenders of the existence of such seemings precisely use this fact to distinguish these seemings from attitudes like belief.[7] Just as it can visually seem to me that something is further away than I rationally judge it to be, so too things can 'intellectually seem' true that one rationally judges false, and for roughly the same reasons. But if that is right, then why does the IMP hold? To judge an inference is bad would just involve judging that the 'intellectual seeming' that accompanies an inference is deceptive. And that should often be rational. One could of course develop a notion of 'intellectual seeming' that does not share its rational profile with ordinary visual seemings, say. But, at best, such an account risks a serious loss of explanatory power by breaking analogies with ordinary seemings and, at worst, threatens to be *ad hoc*.

Because of this, the Intuitional Account faces noteworthy obstacles in accounting for the indefeasibility of the IMP. But another way to run into trouble with the IMP is to make room to accommodate it without actually *explaining* it. For example, Boghossian proposes that inferential-taking involves a form of rule-following behavior that is taken as a kind of primitive. This account hardly *conflicts* with the IMP. For example, Boghossian could propose that one cannot follow a rule and judge it bad to have followed it. But in taking rule-following to be a primitive the account seems to forgo the resources to give a helpful positive account of why this claim holds.[8]

A similar, though less serious, version of the concern faces the account of Wright who, as we've noted before, treats inferring as a form of acting-for-a-reason. Again, as with Boghossian's account, I see nothing in Wright's account to *preclude* an explanation of the IMP. Still, it is worth flagging a tempting avenue for Wright's view that I don't think succeeds. This avenue would appeal to the seemingly general truth that one cannot rationally, and consciously, per-

[6] E.g., Dogramaci (2013), Chudnoff (2014), and Broome (2014)—though Broome only appeals to a 'seeming' as a complementary component to a dispositional account of taking.

[7] E.g., Chudnoff (2013, 44).

[8] For the record, I have great sympathies with the outlines of Boghossian's view. My only issue is with treating the following of rules as a primitive. I think that the account to be given in §5.3 could possibly count as an instance of Boghossian's but for the assumption of non-reduction. I also think there are available, compatible, *broadly* reductive ways of explaining what it is to follow a rule along lines I read into Wittgenstein's later work. See Shaw (2023) for a discussion.

form act A for a reason R while simultaneously judging that R does not provide good reason to A. If inference is an act of accepting for reasons, can't we simply see the IMP as an instance of this more general phenomenon?

Surprisingly, it does not seem we can. The key issue is *what* the reasons involved in inference are reasons *for*. I think the natural elaboration of Wright's proposal is one on which, in inference, one accepts a conclusion for reasons given by premises. But if *this* is how we understand matters, it is not clear that the IMP can be explained in the above manner. The problem is that we will not have a clear account of why the IMP extends to inference under supposition or imagination. For *these states do not generally require epistemic reasons*. I can suppose (especially counterfactually) what I know to be false. I can probably even suppose necessary falsehoods and illogical claims at will—for example, perhaps I am doing so to work out a story to amuse myself. Of course there may be *practical* reasons against supposing in these ways (e.g., again, maybe I should be washing the dishes instead of supposing stories for my own entertainment). But beyond those practical reasons there is very little to rationally constrain an activity like supposing.

Consider a case where I suppose a recognizably necessary falsehood (as I am idly building up a story) but also judge there to be no epistemic grounds for believing it or even supposing it. Perhaps I even judge there are no practical reasons for doing so. Even so, this does not seem deeply incoherent. Even if there is a rational problem (which is not obvious), it seems like it cannot amount to more than a mild form of practical irrationality. The same is true if I first suppose p, then suppose q, and judge that there is no reason of any kind to judge or even suppose q. So the question arises: why if I *infer* q from p and then judge p not to be a good reason for accepting q is there any rational problem? After all, I never needed any epistemic reasons for supposing q in the first place, and could rationally suppose it even if I acknowledged a complete lack of epistemic reasons for doing so.

This is not to say that there is no explanation of the IMP forthcoming on Wright's view. (Indeed, I suspect the explanation I eventually give may be compatible with his framework.) The point is that it not *obvious* how to explain it. Even if inferring is accepting for reasons, it is not clear that a general rational obstacle for judging one has acted on bad reasons can be used to explain the IMP in its full generality, in light of suppositional inference. And if that is not the explanation for Wright, what is?

So, some theories of inference like the Intuitional Account run into trouble accounting for the IMP. Accounts like Boghossian's and Wright's perhaps do better in being compatible with possible explanations of it. But on Boghossian's account we forego important resources to explain the IMP. And even on Wright's account, which has more resources, it is not entirely obvious what shape the account should take.

This can seem to pressure us to adopt views, unlike those surveyed so far, that are tailor-made to account for the IMP: accounts on which inference involves a *belief*, or belief-like state, with a content (roughly) that the inference one performs is a good one. Such accounts of inference, advocated by theorists like Neta (2013) and Valaris (2014), would neatly explain the IMP. To judge that one's inference is not good, while making the inference, would require believing a simple contradiction. Consciously inferring q from p requires believing the inference is good. Simultaneously believing that the inference is not good is rationally incompatible with this presupposed belief, for the same reason that consciously believing any set of transparently contradictory propositions is. So we have a cut and dry explanation of Moorean absurdity.

Despite the attractions of this approach, I am concerned it runs headlong into problems with a very different constraint on an account of inference *also* helpfully brought out by Hlobil. This is that trusted, reliable testimony seems insufficient to position one to perform an inference correctly. But if the 'appreciation' of the goodness of an inference amounts to a belief, it is unclear why this would be so.

Here is Hlobil's example to illustrate the problem. (He borrows the following bit of terminology from Dogramaci (2013): a *hard consequence* of a premise-set for a reasoner is a conclusion the reasoner is not able to infer from the premise-set in a single-step inference.)

> …G. H. Hardy was visiting the mathematical genius Srinivasa Ramanujan…and remarked that he travelled there in a cab with the [dull] number 1729…Ramanujan replied…it was an interesting number because it is the smallest number expressible as the sum of two positive cubes in two different ways…[L]et us assume that Ramanujan made the single-step inference:
>
> (P1) The cab number is 1729.
>
> (C) Therefore, the number of the cab is the smallest number

> expressible as the sum of two positive cubes in two different ways.
>
> Even for a mathematician of Hardy's calibre, this conclusion is a hard consequence of the premise. Only a genius like Ramanujan immediately 'takes' (P1) to support (C)…Now suppose Ramanujan told Hardy that (P1) supports (C). Would this have enabled Hardy to make the very same inference as Ramanujan did? …That seems incredible. Of course, Hardy can make the following inference:
>
> (P1) The cab number is 1729.
>
> (P2) If the cab number is 1729, then the number of the cab is the smallest number expressible as the sum of two positive cubes in two different ways (as Ramanujan just told me).
>
> (C) Therefore, the number of the cab is the smallest number expressible as the sum of two positive cubes in two different ways.
>
> But this is a different inference. It has two premises, whereas the inference that Ramanujan made has only one premise.
>
> (Hlobil, 2019, 7–8, footnotes suppressed)

Roughly, the point is that the appreciation involved in an inference requires 'seeing' a connection between premises and conclusion, and that this ability isn't imparted simply by telling someone that such a connection exists. Even if we *believe* an individual who reports such a connection—indeed even if we do so reliably, and come to *know* the connection exists—this won't be enough for us to make a single-step inference from the premises to the conclusion. It at best gives us a *new premise* to use in another form of inference, as Hlobil notes.[9]

This strikes me as a powerful consideration against views that treat inference as subsuming a belief that premises in the inference support its conclusion. It is of course not a knock-down objection to such views. One can, for

[9] Note: testimony may of course sometimes, somehow empower us to make a single-step inference—intuitively by getting us to 'see connections.' What is striking is that mere acceptance testimony does not *typically* enable us to do so. Indeed, sometimes for extremely hard inferences it *cannot* do so. And this is surprising if all an inference requires is a belief of the sort easily transmitted by testimony.

example, think that the belief required in good inference is a 'special kind' of belief that, unlike others, can't be transmitted by testimony. Or one can think the belief must be held alongside other important conditions for a rational inference to be enabled. Even so, Hlobil's problem of testimony compounds the problems from the IMP. The IMP provides pressure to treat inference as involving something like an implicit belief. But the problem of testimony seems to show some conceptual distance between ordinary belief in the goodness of an inference and the ability to perform it.

However we choose to respond to Hlobil's example, it seems to place a further constraint on an account of good inference: that testimony that an inference is good, and so the belief in or even knowledge of its goodness, are not generally sufficient to enable a thinker to rationally perform the inference in question.

(III) *Insufficiency of Knowledge*: An account of inference should explain why knowledge that an inference is good is generally insufficient to enable one to rationally perform the inference.

While Hlobil's problem of testimony provides evidence against the claim that belief in an inference's goodness is sufficient to rationally perform the inference, it is perhaps worth adding that there is also a problem for the claim that such belief is necessary. This is a point emphasized by Boghossian, among others: if we are willing to grant that young children or even animals engage in inferences, there may be pressure against taking the 'appreciation' of an inference's goodness as requiring the deployment of any sophisticated concepts. Even the concepts of *inference* or *good inference* seem like they would outstrip the cognitive capacities of children and certainly animals.

Now, we should be cautious because many theorists are willing to deny young children or animals the ability to make inferences.[10] Still, I think it should be uncontroversial that younger children can do something that is *like* inferring. Young children can clearly 'draw out' the conclusions of their attitude states (including under suppositions) in ways that are rationally assessable, and amount to more than a mere causal chain of acceptance states. And they can do this without yet appearing to have any sophisticated concepts to classify their reasoning processes. So perhaps the best way to think of the constraints imposed by unsophisticated cognizers is that they require us *either* to

[10] At least in a 'full blooded' sense. See, e.g., Marcus (2021).

make sense of how such reasoners could perform inferences *or* to give a proper account of the continuities between them and sophisticated inferrers that does not collapse their cognitive activities into a single process worthy of the name "inference".

(IV) *Sophistication*: An account of inference should explain how inferring is possible for relatively unsophisticated reasoners like young children, or explain how 'inference-like' activities of such reasoners do not properly count as inferences.

Constraints (II)–(IV) show that a proper account of inference must strike a delicate balance to account for inference's rational role. *Moorean Incoherence* suggests that conscious inference rationally regulates certain kinds of belief. But constraints like *Insufficiency of Knowledge* and *Sophistication* pare back the resources we have to account for that rational regulation.

There is a similar balancing act that arises when it comes to explaining another important role of inference in rationalizing acceptance states. For the reasons laid out in Chapter 3, we need to be careful about how exactly we frame this rationalizing role. For example, the goodness of an inference does not entail it should be performed. Also, inferences mediate between states like suppositions which have nebulous criteria for rational formation (if there are such criteria at all). Even given these caveats, though, it should be acknowledged that inference sometimes plays an indispensable role in mediating between justified beliefs. For example, sometimes when one infers q from a justified belief in p, the belief that q itself counts as justified *because* it was so-inferred. We need some account of how inference can facilitate the transmission of justification and the expansion of our knowledge, at least in certain contexts.

A key problem for explaining this rationalizing role is precisely in accounting for the involvement of an appreciability constraint on inference. Rational inference requires more than simply one belief following another, even causally. So there appears to be something an inference 'adds' to a succession of attitude states that does the rationalizing work. But natural ways of elaborating how this addition plays an indispensable rationalizing role threaten to generate various kinds of regress. After all, what rationalizes the additional element? For example, if the additional element is an appreciation or 'taking,' what makes this appreciation or taking *itself* rational?

This concern has been explored from many different angles in the existing literature. Here I'll mention three.

A first aspect of this worry comes from Boghossian (2014). One of the points Boghossian stresses is that it seems like the 'taking' he pinpoints as an essential component of inferring is itself rationally evaluable. But, he notes, it cannot be that the rationalization of the 'taking' is itself *always* rationalized by inference, otherwise we appear to set off on a vicious regress: each inference must be preceded by another that rationalizes it, which is impossible for a finite mind.

A second worry of this kind traces back to Carroll (1895). There are many different, and I think equally interesting, questions one can take away from Lewis Carroll's parable of Achilles and the Tortoise. The one I want to focus on here concerns *compulsion*. In the dialog, the Tortoise notes that there are several components to a logical inference. Focusing on Modus Ponens, there are the premises: p, and *if* p, *then* q. And there is the entailment relation: q follows from p and *if* p, *then* q. What the Tortoise points out is that each of these components is critical to understanding why a reasoner should draw the conclusion q, when they should. If they don't believe that p, or *if* p, *then* q, then they are under no logical compulsion to draw the conclusion. As the Tortoise puts it, they are not yet "under any logical necessity" to accept it. But there is some sense that the reasoner is also under no such compulsion if they fail to recognize that q follows from p and *if* p, *then* q: recognizing the entailment is also part of what it is to infer correctly, and for the right reasons.

Carroll's dialog then draws out that the role of recognizing the entailment seems different from that of accepting the premises. If the agent merely fails to believe p, then as soon as they do come to believe p as a premise, they come back under the logical compulsion to accept q. But if they fail to recognize the entailment, it is not obvious that believing the entailment as a premise creates the logical pressure to believe q. After all, the reasoner may yet fail to recognize that q follows from p, *if* p *then* q, and *if* p *and (if* p *then* q*),* q. If there was a problem when the reasoner initially failed to see the connections between premise and conclusion, it seems to have persisted. And, familiarly, from here a regress ensues. So the question arises: what is the role of recognizing the entailment, if it is not the same role as that of believed premises? And especially: How can that recognized role sometimes place the reasoner "under [a] logical necessity" to draw the conclusion?

Kripke (forthcoming) puts his own spin on Carrollian regress in what has been come to known (following Padró (2015)) as *the Adoption Problem*. The

concern is that there are certain inference rules that are so fundamental they cannot be 'adopted,' in the sense that one cannot rationally come to accept them in the same ways that we might other truths. Again following Padró, we can say that one adopts a rule of reasoning R just in case:

(i) one does not yet reason with R,

(ii) one comes to accept that R is a good rule, and

(iii) this acceptance *rationally* leads one to reason with R.

The Adoption Problem is that there are a some rules that it seems *impossible* to adopt. Consider the logical rule of Universal Instantiation, or R_{UI}. Suppose some agent, Harry, cannot reason with R_{UI}. So he believes $\forall x(Fx)$, but cannot infer Fa for some a. Can Harry *adopt* R_{UI} as per (i)–(iii) above, and thereby come to accept Fa?

Let "UIForm" be a binary predicate that applies to pairs of sentences just in case they give the form of an inference by R_{UI}, and suppose Harry can apply this predicate just fine. So we tell Harry:

(UI) $(\forall x)(\forall y)(\mathrm{UIForm}(x, y) \wedge \mathrm{True}(x) \rightarrow \mathrm{True}(y))$

And Harry believes us, so he accepts that (UI) is true. He also accepts True("$\forall x(Fx)$") and UIForm("$\forall x(Fx)$", "Fa"). Can he now infer Fa? Well, it seems to use (UI) for that purpose, he would have to get to this instance before applying Modus Ponens.

(Inst) UIForm("$\forall x(Fx)$", "Fa") $\wedge$ True("$\forall x(Fx)$") $\rightarrow$ True("Fa")

But to do *that* he must *first be able to use R_{UI} on (UI)*. And if he can do this, he did not meet condition (i) for adoption in the first place, as he could already reason with R_{UI}. Since Harry was just an arbitrary agent who had yet to reason with R_{UI}, it follows that R_{UI} cannot be adopted.

Kripke notes that this particular form of regress makes it hard to see how rules of reasoning, and in particular fundamental rules of reasoning like those of logic, could just count as 'more beliefs' (a point which obviously dovetails with constraint (III), *Insufficiency of Knowledge*, above). But if these rules of reasoning are not simply more theory, then what does 'accepting' or 'endorsing' these rules amount to?

While the regresses of Carroll, Boghossian, and Kripke are different, they are obviously connected (indeed, both Boghossian an Kripke cite Carroll as an antecedent for their respective regresses). In particular, they can all be viewed as exploring complementary challenges to understanding the rational role of an inference. Both Kripke and Boghossian show ways we are pressured *not* to view how the rationalizing component of an inference is itself rationalized. In particular, it is hard to see this component as itself accepted or reasoned to by inference. But Carroll's formulation of regress also reminds that what we ultimately want from the rationalizing component of inference is quite demanding: it not merely licenses certain attitudes but sometimes rationally forces or compels them. So again we have a balancing act. We need inference to play a strong rationalizing role, when we are stripped of some of the more straightforward methods of supplying it.

Putting these ideas together we have another, composite constraint on a complete account of inference.

(V) *Rationalizing without Regress*: An account of inference should explain the rationalizing feature that distinguishes inference from a mere succession of attitudes. And it must explain how inference rationalizes acceptance states, sometimes compelling them, without creating a problematic form of regress (framed in slightly different ways by Carroll, Boghossian, and Kripke).

My next constraint on inference concerns contingent aspects of its role in the human cognitive economy. As noted above, some inferences are 'hard consequences' of a premise set for a reasoner: the reasoner simply cannot infer to the conclusion from the premises in a single step. Hlobil's work reminds us that the hard consequences for an agent are not overcome by believing or even knowing the consequence relation holds. But we can, I think, go much further than this and note that for most ordinary reasoners, deductive inferences tend to proceed in relatively small steps such as those enshrined in simple logical rules like Modus Ponens, Universal Instantiation, and so on. To be sure, some exceptional reasoners go well beyond this. But it is striking that for ordinary human agents deductive inference proceeds in such small steps. Why? A good account of what makes an agent able to appreciate the goodness of a deductive inference should help explain this.

(VI) *Small Steps*: An account of inference should help explain when, and

> why, single-step inferences are rationally available for a given inferrer—and in particular why deductive inferences tend to proceed in relatively small such steps for ordinary human reasoners.

The foregoing six constraints on an account of inference should make clear why appreciability remains such a challenging and contested topic in the philosophical study of inference. We must say what is 'added' to a chain of acceptance states, beyond causation, that makes them into an inference (*Deviant Chains*) where this added element mysteriously cannot be transmitted by testimony, nor does it seem to amount to something like simple knowledge of an entailment (*Insufficiency of Knowledge*). It must be able to play a strong rational role, in consciously precluding certain acceptance states (*Moorean Incoherence*) and also in rationally compelling others (as per the Carrollian component of *Rationalizing Without Regress*). Furthermore it seems itself to be subject to standards of rational evaluation, even though many means of understanding how the relation could be rationalized lead to regress (*Rationalizing without Regress*). In connection with all this, the relation seems to be cognitively demanding, in that for ordinary humans it can only mediate between small deductive steps (*Small Steps*). But in spite of its cognitively demanding character, there is countervailing pressure to see it as available to even the most cognitively limited reasoners (*Sophistication*). It is often hard to see how anything could satisfy some *pairs* of these constraints, let alone *all* the constraints jointly.

As I've indicated above, there are many existing attempts to navigate this series of constraints, or at least some subset of them. I will not be able to do justice to all of these attempts here. Instead my focus will be on developing my positive account, which will come in the form of an analysis of deductive inference.

5.2 Interlude on Inquisitive States

On the account of deductive inference I will shortly defend, inference reduces to a pair of aspects of cognition. The first of these aspects is the crowding-out states discussed in Chapter 4. The second aspect of cognition will involve forms of *inquisitive states*. Unlike with crowding-out states, I will simply be borrowing a pre-existing, more-or-less received view of how inquisitive states

function. Let me survey aspects of that received view that will matter for the account of inference to come.

The standard view of the semantics of interrogative expressions like *who stole the cookies?* is that such interrogatives semantically express *questions* which are a distinctive form of representation—i.e. distinct from propositional representation. Propositions are typically modeled by a set of worlds where the proposition is true, or by abstract entities structured out of objects, properties, or concepts that determine the conditions under which the proposition is true. Questions, by contrast, are typically modeled by a *partition* of worlds, or a *set* of abstract propositions (for reasons we'll discuss shortly).

In addition to playing the role of the objects semantically expressed by interrogatives, questions are also taken to be the objects of certain cognitive states. Just as states of belief and knowledge take propositions as their objects, there are states like wonder, inquiry, and active suspension of judgment that take questions as *their* objects. The most extensive defense of this idea is found in a series of paper by Jane Friedman,[11] who convincingly argues that inquiry is a fundamentally cognitive matter, irreducible to any set of mere actions or non-attitudinal dispositions and, more specifically, that inquiry is marked by the possession of certain 'interrogative attitudes' like curiosity and wonder which take questions as their objects. Even settled suspension of judgment, which can be a provisional endpoint of inquiry, takes a question as its content in this way.[12] As Friedman notes, although this position requires defense along many dimensions, it receives a tremendous amount of *prima facie* plausibility from a longstanding treatment of interrogative attitude reports in natural language as taking questions as their semantic objects.[13] This literature in some ways makes Friedman's position the default view. So I will not review Friedman's more specialized arguments for the default position here, and will instead simply take it for granted.

Just as there are heated debates about the nature of propositions, there are corresponding debates about how precisely to characterize the semantic values of interrogatives. Some influential positions treat questions as sets of possible

[11] Friedman (2013a,b, 2017a,b).

[12] See especially Friedman (2013a,b).

[13] See, e.g., §4 of Cross & Roelofsen (2020), and the citations therein.

answers,[14] sets of true answers,[15] partitions of logical space,[16] or lambda abstractions.[17] But while there are important differences between the formal semantic objects used to model questions, they are linked by a common underlying idea, helpfully articulated by Friedman:

> One of the main features of nearly any theory of questions is that just as propositions have/are closely related to truth conditions, questions have/are closely related to *answerhood conditions*. This doesn't mean that questions should be answers, but that they should be the sorts of things that can be answered. Moreover, it should be somewhat clear what the conditions under which they will be answered are. The main theories of questions … do this (albeit in different ways). Each account makes a question the sort of thing that to some extent specifies the conditions under which it will be answered, but each also makes the question itself something distinct from those answers.
>
> (Friedman, 2013a, 166–7)

So: roughly, questions specify the conditions under which they are answered. But there are several ways that a question can do this. A question can specify its answers obliquely, roughly by representing a property that the answer will have. Accordingly, when one cognitively relates to such a question, although one is thereby in a representational state, there is no presumption that one thereby represents all, or even any, of the question's answers. For example, consider the question: "What is Hazel thinking about?" There are innumerable possible answers to this question: Hazel is thinking about cookies, Hazel is thinking about paragliding, etc. But an inquirer who is wondering this question needn't be representing all of these answers. There are far too many for a finite mind to encapsulate. What is more, the inquirer needn't even be representing the correct answer. If I'm wondering what Hazel is thinking, perhaps she is thinking a thought involving concepts I do not yet possess. I can still wonder what she is thinking, even if I am not yet in a position to represent the answer to my inquiry.

[14] Hamblin (1973).
[15] Karttunen (1977).
[16] Groenendijk & Stokhof (1984).
[17] Hausser & Zaeffer (1979).

Now, even though questions may represent their answers obliquely, they may equally represent their answers more directly. Two question types are salient in this regard. The first question type is that of *polar questions*—i.e. 'yes/no' questions. Is Hazel thinking about cookies (or not)? In representing this kind of question, you represent its possible answers—the question is such that, semantically, it somehow *subsumes* a representation of those possible answers. The same is true of the second question type of *enumerative questions* like: "is Hazel in her room, or in the kitchen, or in the basement?" Again, semantically, this question does not merely encode a property that answers to that question have, but it encodes the possible answers themselves, directly.[18] Accordingly, cognitively relating to this question, and so representing it, requires representing the question's possible answers directly.

I'll call questions that encode their answers directly *specifying questions*. This will be the kind of question that matters in deductive inference.

In addition to characterizing this special semantic object that matters to deductive inference, I will also need to specify a *type* of state of mind that subsumes both inquisitive states and more familiar states that take propositional objects.

> A is *sensitive to the question* Q just in case A is cognitively engaging with the question of whether Q either
>
> - by bearing an inquisitive attitude to Q, or
> - by having a settled answer to an answer to Q.

Consider three investigators inquiring into a murder and discussing their theories amongst themselves. The first, after reflection, thinks the butler did it. The second thinks the butler *did not* do it. The third thinks the evidence doesn't settle the matter. The butler might have done it or might not. So the third investigator suspends judgment on the issue.

[18] Friedman notes a related point that speakers who understand a polar question know its possible answers: "When it comes to polar questions…there is a kind of "semantic transparency" from questions to answers. Anyone who understands the question will have a good sense of what the possible answers are." (Friedman, 2013a, 159) For the record, although I think polar and enumerative questions give examples of the phenomenon I am interested in, my arguments to follow do not depend on this. Instead they merely depend on the *possibility* of questions whose representation requires representing their possible answers. There could in principle be such questions even if they are never expressed in language.

All three of these characters are sensitive to the question of whether the butler committed the murder. They technically have different 'settled' attitudes states in virtue of which they are doing so. Some (indeed, in some sense, all) of these states conflict. However there is something representational in common between all three characters at a suitable level of abstraction. They are all intuitively engaging with the very same question. The same issue is 'on their minds.' The notion of sensitivity to a question is meant to capture this idea.

With the notion of sensitivity to a question in hand, we have all the resources we need to formulate our reduction of deductive inference, so let me turn to that next.

5.3 A Reduction of Deductive Inference

I propose that a deductive inference is essentially the crowding-out of answers to a question to which one is sensitive. In particular, deductive inference can be given a reductive analysis as follows.

> A deductively infers q from $p_1,\ldots, p_n$ if and only if in A's cognition, crowding-out states are recruited for the purposes of an act of information extraction with the following character:
>
> (i) A is sensitive to the specifying question of whether or not q while accepting $p_1,\ldots, p_n$,
>
> (ii) in representing as per (i), A comes to crowd out a representation of $\neg q$ and $p_1,\ldots, p_n$, and
>
> (iii) A thereby comes to accept q alongside $p_1,\ldots, p_n$.

Recall that "acceptance" extends not only to beliefs, but suppositions and imaginings. I will sometimes, suggestively, rewrite "crowds out not-q and $p_1,\ldots, p_n$" as "appreciates/recognizes/takes it/sees/grasps that q follows from $p_1,\ldots, p_n$" or "appreciates/recognizes/takes it/sees/grasps that $p_1,\ldots, p_n$ entail q". But it is important to remember that these will always merely be notational variations of the first, more fundamental formulation.

All three conditions in the above analysis require clarification.

To say in (i) that A is sensitive to the specifying question of whether q, while accepting $p_1,\ldots, p_n$, is to say that A has engaged, or is engaging, in the

cognitive task of trying to settle whether to accept q or not q, in the context of their accepting propositions p_1,…, p_n. As noted in §5.2 there are many ways we could semantically represent such a question. I will provisionally construe a question as the set of its possible answers. What object of this sort is an agent sensitive to, when they are sensitive to the question of whether or not q *while accepting* further propositions? I mean for this terminology to subsume two possible sets of attitude states. On the first construal, one is sensitive to whether or not q while accepting $p_1, \ldots, p_n$ by being sensitive to a question in which the latter propositions are explicitly represented (e.g., as conjunctions). This would mean one is sensitive to the following question:

$$\{q \wedge p_1 \wedge \ldots \wedge p_n | \neg q \wedge p_1 \wedge \ldots \wedge p_n\}$$

On the second construal, one is sensitive to whether or not q while accepting $p_1, \ldots, p_n$ if the only question represented in one's attitudes is whether or not q, but the premises are also separately accepted. That is, in this case one is sensitive to the following question:

$$\{q | \neg q\}$$

But this question is only engaged with 'alongside' a further acceptance state with content $p_1 \wedge \ldots \wedge p_n$, or several further acceptance states with contents p_1, …, and p_n respectively. Also, since we are stipulating the question is a *specifying* one, we presume that on either of the above construals the possible answers to the question (q alongside the premises, $\neg q$ alongside the premises) are explicitly represented in representing the question itself.

Now, each of these two ways of understanding the sensitivity reported in (i) will come with a corresponding way of understanding what it is for A to come to crowd-out not-q and p_1,…, p_n in (ii). If one is sensitive in the first way, crowding out not-q and p_1,…, p_n will simply amount to crowding out $\neg q \wedge p_1 \wedge \ldots \wedge p_n$. If one is sensitive in the second way, then crowding out not-q and p_1,…, p_n will require crowding out the *collection* of contents $\{\neg q$, $p_1 \wedge \ldots \wedge p_n\}$ or $\{\neg q, p_1, \ldots, p_n\}$. The second form of sensitivity will require that the collection of contents is *integrated* in the sense briefly sketched near the end of Chapter 4.

Note the rider "in representing as per (i)…" in the formulation of (ii). This is not merely indicating temporal conjunction—a co-occurring of the

crowding-out state alongside sensitivity to a question. Crowding-out states, as emphasized in Chapter 4, are not independent states, but are essentially *properties* of pre-existing representational states. In particular, the are modes of a pre-existing representation. The rider in (ii) is indicating that (i) supplies the representational state whose mode is being specified further by (ii).[19]

This brings us to condition (iii), where the critical matter is understanding the force of the "thereby" in saying "A thereby comes to accept q alongside p_1,…, p_n." On the construal I intend, the crowding-out in (ii), as modifying the state given by (i), is not causing acceptance of q. Nor does it otherwise merely characteristically lead to that acceptance (e.g., by rationalizing it). Rather, the crowding out in (ii), in conjunction with the conditions in (i), *metaphysically necessitates* the acceptance in (iii). In fact, for this reason condition (iii) is effectively redundant: it could in principle be left out of the reduction of inference if we liked.

This can be shown by a simple argument. Condition (i) tells us that A is sensitive to a specifying polar question Q. There are three ways to be so-sensitive:

(a) bear an inquisitive attitude to Q,

(b) accept the negative answer to Q, or

(c) accept the positive answer to Q.

(ii) tells us that *in this state* A crowds out the negative answer to Q, thereby precluding a representation of that answer. But being in the states described in each of (a) and (b) requires the capacity to represent a negative answer to Q. This leaves (c)—accepting the positive answer to Q—as the *only* way to remain sensitive to Q while crowding-out its negative answer. So that must be the state that A is thereby in.

We might put this as follows: *in inference, one squeezes an answer out of a question by constraining one's cognition so that only that answer is representable*. Again, on this picture, (i)–(ii) are not causing or otherwise bringing about acceptance of q. They are not (or at least not merely) rationalizing an acceptance of q. They are metaphysically necessitating that acceptance. Another way of

[19] This is integral to ensuring that the 'taking' or inferential step *bind* properly to the premises to drive through an inference without requiring a further intermediary or added cognitive ingredient. I'm grateful to Eric Marcus for pressing me to clarify this issue.

putting this point is to say that, against the backdrop state of inquiry given by (i), grasping an entailment is *a way of coming to accept*—a way of settling the question to which one is sensitive.[20] Alternatively: grasping the entailment is a state *in virtue* of which one accepts the conclusion.

Note, in showing that (i) and (ii) necessitate (iii) we show that an agent who both *maintains* acceptance of the premises and a sensitivity to the issue raised by the conclusion also accepts the conclusion as soon as they see the conclusion follows from the premises. This leaves only one way to avoid accepting the conclusion consistent with grasping the entailment: abandoning acceptance of the premises.[21] This would be a natural way to 'recoil' from a seen entailment, rather than following it to its conclusion. So another way to formulate the foregoing argument would be to say that if one grasps that q follows from $p_1,\ldots, p_n$ in the context of sensitivity to the relevant question bearing on one's acceptances, one either abandons belief in the premises, or one believes the conclusion. Again: I'm so far not claiming that this is rational, or tends to be caused by a state of crowding-out in that context, but that these are the only possible states for the agent to be in.

On the account just given, to deductively infer a conclusion from some premises is to crowd-out a representation of alternatives to that conclusion in light of accepted premises. To crowd-out such representations in that context *just is* to accept the conclusion. What benefits does analyzing deductive inference in this way have?

One set of virtues are those that accrue to constitutivist accounts of inferential appreciation generally—at least on one reasonable use of the label "constitutivism". According to this form of constitutivism, taking premises one accepts to imply a conclusion *is* to accept the conclusion, albeit in a special way.[22] This, as we will see shortly, provides safeguards against *some* forms of regress

[20] In this way, inferring (at least in the doxastic setting) becomes a *species of judging*—a view with a noteworthy antecedent in Frege. See my epigraph of Chapter 1 and Frege (1906, 387).

[21] Could the agent also *merely* cease being sensitive to the question while maintaining an acceptance of the premises? Not while grasping the entailment *in the relevant way*. For recall that this grasping is in fact a *mode* of the representational state given by sensitivity to a question. To abandon the sensitivity is to abandon the representational state of which the grasped entailment is a characterizing feature—and so to abandon the materials out of which that grasp could be constructed.

[22] For different kind of view worthy of the name "constitutivist", see the 'Hereby-Commit' account of inference in Blake-Turner (2022) which roughly inverts the kind of structure I'm attributing to constitutivist views by allowing inferential transitions to constitute taking states.

worry as already appreciated in the constitutivist account of Valaris (2014).[23] As also noted, however, Valaris's account is one which understands inference as requiring a special kind of belief or belief-like state that one's premises support the conclusion. I have concerns that this position faces challenges from Hlobil's problem of testimony discussed in §5.1. Perhaps more importantly, we saw grounds in Chapter 4 to distance crowding-out states from representational states like those of belief or even knowledge. So I cannot help myself to Valaris's particular form of constitutivism and maintain some of the benefits I want from the analysis of inference given above.[24,25]

A constitutivist approach closer to that I favor—so close it is worth taking some time to distinguish the accounts—has recently been put forward independently by Marcus (2020, 2021). Marcus, citing Kimhi (2018) as inspiration, does at least three things. First, he argues that under certain conditions—in which one has what he calls a 'qualifying understanding' of a contradiction—it is impossible to believe it.[26] Second, he aims to draw attention to the fact "that there is a mode of believing such that it is impossible to hold a pair of beliefs $\{p, \neg p\}$ in this mode," though he is careful to flag he does not have an explicit argument for this claim.[27] And finally, he suggests that we can see something analogous happens in inference (where again, we are meant to recognize this for ourselves as opposed to accepting it on the basis of argument):

> I have no argument that there is a mode of belief such that it is im-

[23] It is important to note that Valaris distinguishes between 'basic' and 'non-basic' reasoning, and that he only takes the latter to be governed by something like Taking Condition. When I discuss Valaris's views, I will be focusing on his account of non-basic reasoning and inference.

[24] In recent work, Valaris (2020) clarifies that his view of taking states is one on which they are like belief in that they are "representational states with intentional content, and are moreover subject to epistemic evaluation." It should be clear that I do not think of crowding-out relations as involving anything like a representational state with intentional content. I've been at pains to emphasize that it is not a *separate* state of representation, but a representational mode. I have also flagged that it is a delicate matter how we evaluate crowding-out relations, since we can only do this relative to some purpose.

[25] It is also worth noting that it is unclear that Valaris is a constitutivist in the strongest sense—a sense in which both Marcus's view (to be discussed presently) and my view would qualify. In particular, Valaris denies that accepting premises and 'taking' the premises to support a conclusion metaphysically necessitate an acceptance of a conclusion. This can be blocked for him by irrationality or inattention. This is a point which seems to weaken the explanatory power of the theory. See related criticisms in Marcus (2021, §4.5).

[26] Marcus (2020, 5).

[27] Marcus (2020, 6-7).

> possible to hold a belief and its negation in mind together—that is a datum I'm simply counting on the reader to recognize. Similarly, I take the following to be a familiar phenomenon: one has in mind the beliefs that p and that $p \to q$, and so it is impossible not to believe q. There are circumstances in which beliefs of this form possess mental togetherness—there is no inattention, repression, etc. that would explain the failure to draw the conclusion—and so the subject is compelled by her own understanding to accept a conclusion that she recognizes as following logically from the premises that [she] accepts. Under these circumstances—absent the relinquishing of a premise—it is *impossible* for her not to believe the conclusion.
>
> (Marcus, 2020, 9)

Marcus never goes so far as to *identify* the aspects of cognition involved in inference with those that preclude belief in contradictions, as I have done. But it is clear that he takes the phenomena to be closely related. And to that extent, it is clear the resulting view is very close to my own.

In spite of its similarities, here I want to indicate two places where my account diverges from that Marcus gives, each of which are important to its plausibility. The first concerns how Marcus approaches the phenomenon of impossible belief in a contradiction. As I say, Marcus tries to supply an argument for this claim that adverts to the notion of a qualifying understanding of a contradiction (a notion we won't need to probe further to understand my criticisms). The argument runs as follows:

P1: If S believes q, then S has a qualifying understanding of q.

P2: If S understands q to be a contradiction, then S takes q to be false.

P3: If S takes q to be false, then, necessarily, S doesn't believe q.

P4: If S has a qualifying understanding of $(p \wedge \neg p)$, S understands it to be a contradiction.

C: S cannot believe $(p \wedge \neg p)$.

I think this argument is problematic. (P3) begs the question. If an agent takes q to be false, this is very similar to taking its negation $\neg q$ to be true. In this

context (P_3) says that I do not (really the argument requires: cannot) believe things whose negations I believe. Alternatively: I cannot believe the negations of the things I believe. But if I suspect that one can believe contradictions (of which I have a qualifying understanding), surely I will be doubtful of the claim that I cannot believe the negations of things I believe.

The argument is also problematic because it fails to get to the core of the cognitive resistance created by contradictions. As I stressed in Chapter 4, this resistance manifests itself even for supposition or imagination. But an argument like Marcus's does not seem to offer any insight into why this would hold. For example, a transposition of (P_3) to the case of supposition would claim that we necessarily do not suppose what we take to be false. But of course, we can and regularly do make suppositions of this kind with no trouble.

Note that defending a resistance in entertaining impossibilities for all attitudes, including those like supposition which are not rationally regulated by concern for truth, is critical if one wants to pursue a *reduction* of deductive inferential appreciation to something like crowding-out states. This is because the argument I've given above that crowding-out states *necessitate* the acceptance of a conclusion critically depends on crowding-out relations ruling out the possibility of question-taking attitudes like inquisitive states. If crowding-out relations only affect beliefs, it is not clear how they can generate positive acceptance of a conclusion.

This serves as a segue to my second point. As noted, Marcus does not go so far as to identify the operations of inference with the resistances to entertaining certain impossibilities, instead noting key analogies between them. But there are costs to refraining from the identification. If we treat these as different phenomena, we end up positing a plurality of *sui generis* cognitive relations. This is not merely important for reasons of theoretical parsimony and unity. Rather there is a special problem for any aspiring constituvist view of saying *why* the taking or appreciation involved in an inference metaphysically necessitates the drawing of the conclusion. Marcus seems forced without argument to simply state the constitutivist thesis holds. Without supplementation, the analogy with the cognitive resistance from contradictions only seems to exacerbate worries about understanding the necessary force in drawing a conclusion, since that resistance merely *prevents* the formation of certain attitudes. How can this resistance, or even something analogous to it, positively *generate* and *necessitate* an acceptance state?

This is why I think it is important to see an account of deductive inference as growing from a combination of two theories: (i) a theory of how impossibility impedes cognition along with (ii) a broader theory of inquiry, including question-sensitive attitudes. This combination allows us to *demonstrate* how an impediment to cognition can, in a certain context, drive the existence of certain positive attitude states. This occurs precisely because maintaining background states of inquiry require some kind of attitudinal representational state, and crowding-out states render impossible all that don't involve acceptance of the conclusion. This explains the necessity of accepting a conclusion without forcing us to posit any controversial cognitive relations beyond those we needed to account of the cognitive resistance to entertaining contradictions. And we have independent empirical evidence for that cognitive relation.[28,29]

The foregoing discussion aimed to highlight some differences between Marcus's views and my own. But the differences should not be overstated. Our views are closely connected, and many of the virtues I will claim below for my view will accrue to Marcus's as well. The next step is to see what these virtues are, by turning back to reconsider the conditions placed on an account of deductive inference from §5.1. I'll do this now, albeit in a new order.

(I) *Deviant Chains: An account of inference should illuminate why mere causation among acceptance states is insufficient for inference to take place, and describe what connection between acceptance states* is *required for it.*

The reduction of inference that I've proposed has a simple way of explaining why causation is insufficient for inference while avoiding deviant chains. It accomplishes this merely by eschewing causation in the

[28] I've focused on Marcus's early formulation of his views in Marcus (2020). In the more developed work of Marcus (2021), he drops his attempt to provide a deductive argument for the resistance to representing contradictions. But he continues to focus on *belief* as the central case. For the reasons above, I think is both independently misleading (as the core resistance to cognition has little to do with belief in particular, and is rather a feature of representation more generally), and continues to thwart an adequate generalization that would facilitate the form of reduction that I advocate and find especially attractive.

[29] I also take this to be an advantage of the view over other formulations of constitutivism, like that of Valaris (2014, 2020), since these also do not provide arguments that *show* how the key ingredient in inference—in Valaris's case a form of belief—necessitate the drawing of a conclusion. As far as I can tell, he simply posits that it does, seeing as this would help resolve problems of regress.

account of inference at all. It does *that* by pinpointing the rationalizing element in inference as crowding-out states and then simply *equating* deductive inferring with the situated presence of this element. This way of avoiding problems with deviant chains may be surprising, because pinpointing the right notion of deductive taking and resolving deviant causal worries may seem disconnected. In recommending Taking Condition, for example, Boghossian is careful to hedge: "[The problem of deviant causal chains] is still with us [once we adopt Taking Condition]: the 'taking' on which I am insisting has to cause the conclusion 'in the right way.'"[30] But on the current proposal this is not true. We do not need to inquire about the right way for deductive taking to cause the concluding attitude of an inference, because deductive taking doesn't cause that attitude at all. Against the right background, it constitutes the inference. It is impossible for appreciation to take place against the relevant background attitudes I have specified without an inference having thereby taken place, and the conclusion having been accepted. Here we encounter one of the key virtues of the constitutivist line I propose.

(V) *Rationalizing without Regress: An account of inference should explain the rationalizing feature that distinguishes inference from a mere succession of attitudes. And it must explain how it rationalizes acceptance states, sometimes compelling them, without creating a problematic form of regress (framed in slightly different ways by Carroll, Boghossian, and Kripke).*

Why is an inference rationally evaluable, and how can it rationalize? Recall that in Chapter 4 I stressed that relations of crowding-out are not evaluable on their own, independently of some purpose or end read into the representational modes that underlie them. Moreover it is not clear that, considered purely and in isolation, our representational modes have an overriding purpose of adequately modeling metaphysical modal space: that is just one purpose among many that a representational scheme could serve.

This is where it matters that "crowding-out states are recruited for

[30] Boghossian (2014, 5, n.2).

the purposes of an act of information extraction"—a component of the analysis of inference I've offered which ties the reduction back to the functionalist skeletal account of Chapter 2. Recall that on this construal, inference is conceptualized as having a particular role in our cognitive economy: that of reliably extracting information from information-bearing states. My analysis proposes that cognitive states of crowding-out are recruited for these purposes in cognition. It is easy to see why they would be recruited in these ways. For as just argued, they can have *exactly* the features needed to establish relations of informational containment between accepted premises and a considered conclusion.

Once crowding-out states are conceptualized in these terms, they bear two features that make sense of their rational, epistemic evaluability. First, these states serve an epistemic end in a cognitive activity—the end of information extraction. Second, the states are internally subject to modifications (switches in representational modes between more or less perspicuous ones, adopting modes that 'integrate' multiple representations, etc.) which can worsen or improve the ability of the relevant states to achieve their end of crowding out only metaphysical impossibilities to extract information about what is or might be.[31] The first feature situates relevant crowding-out states squarely within the epistemic functions of an agent. The second feature distinguishes their function from those belonging to faculties (like perception and memory) which are arguably not epistemically evaluable as rational or irrational owing to their 'passive' character. A perception in which we experience a persistent illusion (like the Müller-Lyer) is neither irrational or rational of itself, seemingly because the deliverances of our perceptual faculties are not subject to internal modification (as is witnessed by their unresponsiveness to countervailing rational pressure, such as from a stable judgment that the perception is illusory).[32] In this one respect, representational

[31] Note: "only", not "all and only". To achieve its function in ensuring information-preservation in driving through an acceptance of a conclusion q by precluding out a representation of $p \wedge \neg q$, it is enough that $p \wedge \neg q$ be impossible. If other impossibilities are 'compatible' with the representational mode, this will have no bearing on *this* transition being information-preserving.

[32] This claim about perception and memory is contestable. See especially Siegel (2017), who argues that perception itself can be rationally evaluable. Without getting into details, even

modes and the crowding-out relations they underly are unlike perception and memory, and more similar to representational states like belief, for being in-principle subject to such internal modification.[33]

Not only can we make sense of why crowding-out states would be epistemically evaluable, but we can see that they are evaluable in just the terms we would like. The work demanded of our representational modes is that they facilitate an act of information extraction, so this sets the standard for the modes to achieve their end. And representational modes do this precisely by being adequate, in the sense of crowding-out only metaphysical impossibilities. An inference is good, and achieves its cognitive end, if it extracts information via adequate, and so rational crowding-out states. And conversely, one way an inference can be bad is for failing to achieve that end owing to the presence of inadequate, irrational states.[34]

The account on offer can also give us some resources to say how crowding-out states can at least sometimes play a rationalizing role. Inference is a way of settling deliberation through an act of information extraction that is now seen to be epistemically assessable based on that end. Sometimes this cognitive function could be exploited in extracting information from acceptance states, where the extraction of information reliably preserves relationships of rational support.

For example, suppose an evidential support relation for proposi-

if perception were rationally evaluable, the grounds for this would likely just make it *easier* to argue that crowding-out states are epistemically rationally evaluable as well. So I set this view aside here.

[33] Thanks to Ulf Hlobil for helping me see the importance of this condition.

[34] I am tempted to equate the rationality of a crowding-out state with its adequacy, as I do here. This equation of a form of 'correctness' with rationality would not make sense for other states (e.g. we would certainly not want to equate rational belief with correct belief). But the equation makes sense for crowding-out states in part because of their *fundamentality*. They do not have the appropriate structure to be justified or based on anything further—see just below for further discussion. However, even if we *can* equate in these ways, we may eventually want to leave room for a distinction between rationality and correctness for fundamental states of this kind. Perhaps, e.g., someone can rationally crowd-out various metaphysical possibilities in inference when under the influence of certain kinds of misleading tutelage. I won't delve further into this issue here, and merely note it is an interesting choice point worthy of further investigation.

tions transmits through an entailment—that is, evidence for premises $p_1, \ldots, p_n$ just is evidence for q in virtue of the fact that the premises necessitate q. Suppose further that this assumption holds in the context of an agent's formation of new beliefs, given their old beliefs. In this context, a well-performed act of information extraction will be a rationally evaluable way of arriving at a new belief that tracks the relation that transmits evidential support. So if it was rational to simultaneously believe the premises $p_1, \ldots, p_n$ in light of the evidence, it will be rational to arrive the belief in q via this rationally evaluable process which reliably preserves relations of evidential support.

This tells us why inference is rationally assessable and gives *some* resources for understanding how it can rationalize.[35] But has the account staved off regress in the process?

There is no special worry that the rationalizing force provided by crowding-out states might always have to be the result of inference—the concern that Boghossian emphasizes. This is simply because an inference mediates between acceptance states, and crowding-out states are not representational states of *any kind*, let alone representational states of acceptance. One can only arrive at a new representation, not a new representational mode, through an inference. So inferential regress in particular cannot get off the ground.

But it seems like we can say something much stronger. Representational modes are not the sort of thing that can be *justified* on the basis of something else (a belief, a further representational mode, etc.) *at all*. They simply don't have the structure to be 'supported' by anything else. Despite this, they are still evaluable. This is because the use of a representational mode in inference constitutes an activity or process conceptualized under a given end—and that end allows us to evaluate the crowding-out relations which underlie the given representational mode. Inference's rational credentials in this context come from its being adequate or not—from serving its cognitive end well or not—and that is

[35] It is admittedly not a full account of course. Minimally, we would want inference not merely to rationalize, but to generate epistemic basing relations. See n.42 of this chapter for some discussion.

all. This combination makes representational modes ideal to block any form of rational regress: they are evaluable (in the context of an inference) as good or bad, employed well or poorly, but are still not the sorts of things for which questions of support make sense. They inherently form a kind of rational bedrock.

The rationalizing role of crowding-out states in inference is accordingly also compatible with Kripke and Padró's lesson that certain basic forms of reasoning cannot be 'adopted'—that is, taken on board as more theory, justified through the acquisition of new beliefs. The reason, again, is that crowding-out states are neither themselves acceptance states, nor are they justifiable on the basis of acceptance states. So new attitudes can neither constitute nor rationally support the key states needed to facilitate deduction.

Finally, in spite of not being an acceptance state or justified through acceptance states, inference plays exactly the strong rationalizing role that Carroll carves out for it: it *forces* the drawing of a conclusion for a reasoner in the right setting. But it does so while respecting the distinctive challenges put forward by the Tortoise in Carroll's dialog. The Tortoise asks Achilles to "force [him], logically, to accept [the conclusion]" as someone who initially lacks a recognition of the entailment relation. If that is the task set to Achilles, it's true that Achilles cannot fulfill the request merely through trusted testimony, or any other way of imparting new knowledge to the Tortoise. (This is the problem with having Achilles simply write down the entailment—the 'missing element'—in his notebook.) But that doesn't mean the task can't be accomplished. What Achilles must do, if his interlocutor is at all genuine, is to get the Tortoise to 'see' the entailment in the way we've described crowding-out relations to operate. He must get the Tortoise to represent the premises and conclusion in a new, more perspicuous way, on which their logical connections manifest themselves through those representational acts. There is no conventional speech act that imparts that kind of representational mode. And there is no guarantee that, for any agent, it can otherwise be imparted. But the important point is that once it is imparted, the Tortoise will finally have what he alleges to lack. And no stubbornness, recalcitrance, or obtuseness, no matter how gross, could prevent

him from being compelled to draw the conclusion in precisely the way he demands.

This last virtue of explaining how the appreciation in good inference forces the drawing of its conclusion would belong to many constitutivist accounts. But we improve upon accounts like those of Valaris, by separating out the constitutivist move from the claim that what underlies inference is a representational taking state. The latter claim seems to exacerbate concerns about rational regress (as it at least makes sense to ask how this representational state could be justified). This is not to mention that it makes it much harder to account for further key constraints on deductive inference, which we will continue to see below. And we also improve upon constitutivist accounts like those of Marcus *or* Valaris by showing how the necessitating force of an inferential appreciation can be *derived* from the behavior of independently motivated relations of crowding-out, and so explained by them.

(II) *Moorean Incoherence: An account of inference should explain the IMP—that is, how a rational tension generally arises between any conscious inference and the conscious judgment that the inference is a bad one.*

Why is it impossible or incoherent to 'simultaneously' make a conscious deductive inference and consciously judge that the premises don't entail the conclusion? It turns out that the hard work here is done by showing that crowding-out states are epistemically evaluable in the context of an inference (which is why we had to consider *Rationalizing without Regress* first). For once we show that, we can see that Moorean inferential incoherence is an instance of a more general phenomenon.

> Rational Sensitivity to Assessment: In a rationally coherent mind, a conscious epistemically evaluable process or state that is rationally judged epistemically deficient will in typical conditions yield to the judgment.

Rational Sensitivity to Assessment entails that in typical conditions, it is irrational to judge of a sustained epistemically evaluable process or state that it is deficient. For either the judgement itself is irra-

tional, or the judgment is rational and by Rational Sensitivity to Assessment the sustained process or state exhibits irrationality.

We can see the operation of this principle in discussions of what is sometimes called *epistemic akrasia*,[36] which would occur if a subject were to judge something like *p and this judgment that p is irrational* or *p and this judgment that p is completely unsupported by the evidence*, etc. Familiarly there are not only competing accounts of what would go wrong in cases of epistemic akrasia, but differing accounts of how to characterize the phenomenon itself. Some take the combination of attitudes to be impossible.[37] Others take it to be possible but unavoidably irrational.[38] Still others take it to be irrational except in specific, constrained circumstances (my hedge "typical" in the above principle is meant to create room for these theorists).[39]

Which of these accounts of the nature of epistemic akrasia is correct will not be of much concern here, as we can extend any one of them to the inferential case. For example, some claim epistemic akrasia is not possible because cases where one *seems* epistemically akratic don't actually involve genuine beliefs both that p and that p is irrational. Perhaps the apparent belief that p is something more like a brute disposition. Of course, we are free to treat inferential cases along similar lines—with the inference itself being something more brute and dispositional in cases of the IMP.

There are also those who think that epistemic akrasia can be rational in constrained cases. For example, some think it could be rational to believe an expert about epistemology that skepticism is true, and so rational to believe that one is unjustified in believing that one has hands, while simultaneously believing that one has hands (because one has no other rational way of going about things). One can of course imagine similar cases where one trusts an expert in logic that an inference rule is invalid, but one is stuck going around using it because one has no reasonable

[36] See Owens (2002) for an early description of the relevant epistemic phenomenon as a form of akrasia.

[37] Hurley (1989), Pettit & Smith (1996), Adler (1999, 2002), Raz (2009).

[38] Scanlon (1998) seems to talk this way. See also Greco (2014).

[39] Weatherson (2008), Coates (2012).

alternatives. The treatment of these cases seems analogous.

Accordingly, the key point will be the idea that whatever the phenomenon of epistemic akrasia amounts to, what explains it is something like Rational Sensitivity to Assessment. As long as we can understand how inference itself involves an epistemically evaluable process or state, we can see inferential Moorean phenomena as just one more instance of the rational sensitivity of such processes or states to judgements about their rationality. This will be true whatever explains Rational Sensitivity to Assessment (or even if the principle has no informative explanation to begin with).

This may seem like a dissatisfying account of the IMP without delving further into the details of the more general phenomena of which inference is a part. But even if there is room to go deeper by probing the further sources of Rational Sensitivity to Assessment, it is already substantial progress to subsume inferential Moorean absurdity and epistemic akrasia under the same general rational error in the way I have done. In particular, doing so immediately helps us understand the problems with several rival routes to explaining the inferential absurdity.

For example, as discussed in §5.1, one tempting route to account for the IMP on the Intuitional Account of inferring is to say that it must be irrational to judge that a consciously intuited entailment is faulty. The problem is that intuitions are not rationally assessable in the right way to trigger Rational Sensitivity to Assessment: agents are not generally epistemically blameworthy for how things seem to them. In connection with this, it does not take an especially aberrant case for a persisting intuitional seeming to represent something false. Both of these facts explain why it is often perfectly rational to judge of a persisting intuitional seeming that it is a seeming of what is not the case.

Consider also the tempting possible explanation of Moorean absurdity for defenders of an account like that of Wright, on which inferring is just an instance of the more general phenomenon of doing something for a reason. As noted in §5.1, it might be enticing to try to suggest that the IMP is just an instance of a more general phenomenon of judging that

one has no good reason for doing something that is done for reasons. I have broad sympathies with the ideas underlying this proposal. But we've seen that there are certainly ways to apply it that go astray. For it is tempting to think that *what* is being done for a reason in an inference is *accepting a conclusion*, and that the reason provided for it is the reason *given by the premises*. But if we think the IMP results from accepting the conclusion for the wrong reasons, we will be unable to properly account for suppositional reasoning for the reasons I stressed in §5.1: even manifestly false suppositions *need* no reasons to be held rationally.

Indeed, something like this problem arises for an account of inferential Moorean phenomena suggested by McHugh & Way (2018). McHugh and Way frame their view within a more general account of reasoning, on which it is a functional kind regulated by the aim of getting 'fitting attitudes.' The account of inferential Moorean absurdity in the theoretical case goes as follows:

> Theoretical reasoning is guided by the aim of acquiring fitting beliefs. If p does not support q, then reasoning from p to q is not a good way to pursue this aim. So, reasoning from p to q while judging that p does not support q amounts to taking what you acknowledge to be an unreliable means to your end. That looks plainly irrational. [...T]his seems enough [...] to explain why assertions of ["r, so, p; but r does not support p"] seem incoherent.
>
> (McHugh & Way, 2018, 191)

The account suggests that the problem in the IMP lies in one's judging that one has failed to reliably assure the 'fittingness' of the conclusion q. But how can we use this account to explain Moorean absurdity for inference under supposition? As I've stressed, I can suppose virtually anything I want, for virtually any reason, provided I am not otherwise occupied. So in what sense has the 'aim' of reasoning in arriving at fitting attitudes been thwarted in any significant sense if I arrive at the supposition that q on the basis of faulty reasoning? After all, I could simply suppose q now, on the basis of *no* reasoning, without *any* reasons for supposing q, and nonetheless do so without irrationality.

One could say that q is not 'fitting in light of' the supposed premises. But even this doesn't get to the heart of the matter *unless one mentions inference itself* as the locus of rational assessment. After all, I can suppose p, then suppose q (perhaps conjoining it), again without any rational fault. In this context there is no problem with the fittingness of q even 'in light of p.' It matters that it is the *connection* that is evaluated, and the evaluability of the connection is distinct from the evaluability (the fittingness, should there be any) of the attitudes involved. Indeed, for these reasons I suspect McHugh and Way's more general account of the aim of reasoning itself must also be flawed, at least if it is meant to produce anything like a standard of goodness for *individual* acts of reasoning. Inferring under supposition is a component act of reasoning. But it is in no way regulated by an aim of getting fitting suppositions—whatever that would mean.[40]

The problem with McHugh and Way's account is in fact the selfsame problem that we saw beset accounts of the normativity of logic in Chapter 3. This is the problem of mis-locating the norms governing reasoning with the norms governing the attitude states that reasoning mediates between. Even if all acceptance states were generally governed by norms, which is not always clear, we could not work backwards from the norms governing them to the norms governing individual activities of reasoning themselves. Again, this would be like the assumption that the goodness of hammers must lie in the production of good nails, rather than in satisfying an aim specific to the function of hammers.

So I am not the only theorist who has appealed to an explanation using something like Rational Sensitivity to Assessment in accounting for inferential Moorean phenomena. But what we now see is critical to a satisfying explanation of that shape is getting a grip on how and why

[40] For the record, I am sympathetic to the idea that reasoning (construed as a process which subsumes inference as a proper part) is a goodness-fixing kind. It is just, as explained in Chapter 3, that I think that reasoning involves many distinct components, each of which may be evaluable independently of their contributions to the goals of good reasoning. In this way, the relationship of the goodness of reasoning to the goodness of inference is like the relationship of playing the position of pitcher in baseball well to pitching fastballs well. An inference may be good *qua* inference but bad *qua* act of reasoning, just as a fastball pitch may be good *qua* fastball pitch, but bad *qua* activity as a pitcher on mound, in the given context of a particular baseball game.

inference itself is subject to a proprietary standard of epistemic evaluation. That is one thing that the account I've put on offer can provide. And it is non-trivial to provide precisely because of regress worries. But once we have a satisfying account of this kind, Rational Sensitivity to Assessment can be wheeled in to yield a plausible account of Moorean phenomena.

Note the account also explains why *conscious* inference would matter. One is not epistemically akratic for having an unconscious belief that p, and some further (perhaps conscious) belief to the effect that beliefs that p are unsupported. In this case, it is not clear there is enough of a connection between the higher-order belief and first-order belief for the rational pressure exerted by the former to influence the latter. Whatever we say about this case is something we can, and should, say about unconscious inferences as well.

(III) *Insufficiency of Knowledge: An account of inference should explain why knowledge that an inference is good is generally insufficient to enable one to rationally perform the inference.*

As I stressed in Chapter 4, crowding-out states are not a representational states with intentional content, nor is their possession entailed by belief or even knowledge that p is impossible. To crowd-out p is to represent with clarity and precision in a way that precludes the representation of p. Perhaps if p is actually impossible, there may be some guarantee that there is some such way of representing clearly. But even in such cases, to know that p is impossible would at best allow one to know that there is *some* such way of representing. This neither characterizes the *means* by which the mode of representation is arrived at nor, even if it did, would it by itself *enable* a capacity to represent in that way. So we should not expect trusted testimony about an entailment to of itself facilitate the means for good inference. And neither, for the same reasons, should we expect this of knowledge of an entailment more generally.

(IV) *Sophistication: An account of inference should explain how inferring is possible for relatively unsophisticated reasoners like young children, or ex-*

plain how 'inference-like' activities of such reasoners do not properly count as inferences.

In Chapter 4, I stressed not only that relevant forms of belief or knowledge are insufficient for entering crowding-out states, but also that they are *not necessary* for it. We can see this in the example of what precludes the ability to imagine a square circle. The belief or knowledge that such a figure is impossible minimally seems to require the sophistication to think about metaphysical modality. But one does not need capacities beyond those required to represent squares, circles, or figures more generally to represent in a way that crowds out the representation of a square circle in imagination. Rather, one only needs an ability, or facility, in representing plane figures with a requisite degree of precision or clarity. So if an inference is constituted in part by a relation of crowding-out, there is nothing that precludes animals or children from inferring: to infer from p to q, one *only* needs the sophistication to have sufficiently clear, precise, and integrated representations of p and q themselves.[41] I see this as an advantage over accounts like those of Marcus, which frame inference in terms which by his own admission cannot extend in a full-blooded sense to animals and perhaps even young children.

Avoiding a treatment of crowding-out states as separate representational states with their own intentional content thus explains why the conditions on good inference seem to be both strong and weak at the same time: so strong that no knowledge is sufficient to provide it, so weak that even animals or young children could in-principle have it. It is because it is simply a different *kind* of thing from a representational state—instead a *mode* of such a state—that it can simultaneously have these features.

(VI) *Small Steps: An account of inference should help explain when, and why, single-step inferences are rationally available for a given inferrer—and in particular why deductive inferences tend to proceed in relatively small such steps for ordinary human reasoners.*

[41] Though I don't have the space to discuss it here, this feature of the view might qualify it as a neo-Wittgensteinean or neo-Tractarian account of inference, insofar as it vindicates Wittgenstein's claim that inferential relations are 'internal,' and so not mediated by anything external to the propositions themselves (see Wittgenstein (1922, §5.131)). See especially Nir (2021) for an explanation of the stringency of this requirement and how it set Wittgenstein's views apart from those of his contemporaries.

> In Chapter 4, I explored how the empirical evidence from cognitive resistance not only motivates the existence of representational modes that foreclose certain kinds of cognition, but seems to indicate that in human subjects those modes of representation seem to be extremely cognitively demanding. Only for the most transparent forms of impossibility does a typical human reasoner encounter this resistance. If relations of crowding-out figure in deductive inference in the way I've suggested, this leads to the prediction that deductive inference will, in human subjects, proceed in commensurately small steps. And this seems to be just what we find.

The constraints enumerated in §5.1 are hardly exhaustive.[42] But they certainly form a representative set of constraints which are non-trivial to satisfy jointly. What we have seen is that, for each constraint given, the reduction of inference to suitable crowding-out relations either satisfies the constraint outright or provides productive and promising inroads for meeting it. The account simultaneously explains why and how deductive inference is epistemically evaluable, why it has various strong forms of rational force, why it is divorced from belief or knowledge of entailment, why even unsophisticated cognizers can engage in it, and why it proceeds in small steps for ordinary humans. As I've tried to flag, many views account for some of these constraints only at the expense of abandoning others, or explaining them merely through analogy. Even constitutivist views, of my which my account is an instance, struggle with some of these points.

Even if another theory did well in accounting for these phenomena, I would be tempted to lean on an attractive feature of the specifically reductivist aspects of the proposal. As mentioned at the beginning of this section, condition (iii) in my reductive analysis of inference is strictly speaking superfluous. As such, the analysis really only has two components: a crowding-out state

[42] The most significant omission is perhaps that inference characteristically generates 'epistemic basing relations' (see Korcz (2021) for an overview of the phenomenon). My temptation is to see epistemic basic relations formed in inferential cases as states that act as the representational residues of inferences. My analysis of inference describes the *onset* of special set of representational states and modes. But one could tinker with the analysis very slightly (e.g. replace "comes to crowd out" with "crowds out" and "thereby comes to accept" with "in this way accepts") to arrive at a characterization of a standing state—one that needn't always be the product of an inference. I will admit this is only to gesture at first steps, and the topic would have to be dealt with in much greater detail to fully justify the overarching theory of inference on offer. This is beyond my abilities at present and I regret I must leave it for further investigation.

and the state of sensitivity to a question. It is worth emphasizing that these two aspects of cognition are arguably ones *any* theorist will have to accommodate independently. The existence of inquisitive attitudes with the structure I've posited represents a rare point of confluence in theorizing in semantics and the philosophy of mind. And we have millennia worth of philosophical theorizing and now some empirical data that motivate the existence of something like crowding-out relations, even if they are not characterized in exactly the terms I favor. What we are seeing is that we already have two extremely well motivated aspects of cognition that trivially combine to do the mental work we would want of a deductive inference: mentally extracting information from an information-bearing state. Even if something further *could* do this work, why wouldn't the mind to make use of the resources it already has to accomplish this essential cognitive task?

For all these reasons, I think that the reduction I've put on offer articulates an attractive account of the mental activity of deductive inference. That said the account on offer is so far *merely* one for deduction. My focus on deduction derives from my interest in deductive logic. But it is not clear that there can be a satisfying account of deductive inference which does not say *something* about its ties to ampliative inference. In the next chapter, I'll explain why drawing these ties is especially important for my account, and try to make some first steps toward doing so.

CHAPTER 6

Ampliative Inference

In Chapter 5, I focused on accounting for the nature of deductive inference because of its distinctive ties to deductive logic. But any account of deductive inference should be integrable with a more general view subsuming ampliative inference because of both the similarities and the differences between them. For example, some of the constraints I gave on deductive inference in Chapter 5 seem to extend to the case of ampliative inference straightforwardly. These would include *Deviant Chains*, *Moorean Incoherence*, *Sophistication*, and *Rationalizing without Regress*. But it is not obvious that *all* constraints carry over. It is much less clear, for example, that *Insufficiency of Knowledge* and *Small Steps* as formulated apply to ampliative inference.

Eventually we will need an account of these similarities and dissimilarities. It would be a serious concern if the way a theory accounted for constraints on deductive inference did not carry over to ampliative inference when those constraints were shared. It would be equally problematic if the account of deduction did not make room to explain the differences between deductive and ampliative inference.

Indeed, this very kind of worry has recently been raised precisely against constitutivist accounts of inference by Blake-Turner (ms./2021). Focusing on Valaris's account, Blake-Turner claims that the virtues of constitutivism accrue to an undue focus on deduction. In particular, the concern is that Valaris's constitutivism gains its plausibility from considering only cases where premises are 'taken' to *decisively* support their conclusions. Obviously in ampliative inference this does not hold. Blake-Turner argues that Valaris's view is unable to account for appreciability or taking relations in ampliative inference. If true, this is a serious problem: once we account for appreciability in the context of ampliative inference in non-constitutivist terms, wouldn't the account gener-

alize back to the deductive case to render the constitutivist resources superfluous?

For these reasons, no account of deduction—and especially not one which appeals to a form of constitutivism like mine—can get away without saying *something* about the relationship between deductive and ampliative inference. Without delving deeply into these matters, my goal in this chapter is to say enough about the connection between these two forms of inference to show that the account of Chapter 5 is suitably generalizable.

I will begin in §6.1 by giving a limited account of ampliative inference on which it can be viewed as *subsuming* a (sometimes trivial) deductive component. On this view, we can see ampliative inference as a generalization of deductive inference. Alternatively, we can with equal right see deductive inference as a limiting case of ampliative inference. In effect, there is just one mental activity—inference—with two aspects present in varying degrees. In §6.2, I explore how this account generalizes important lessons about deductive inference from Chapter 5, while respecting important dissimilarities between 'pure' deduction and ampliative inference. I conclude by noting several important matters the account leaves to future research.

6.1 Presupposition and its Role in Ampliative Inference

In the account of Chapter 2, I suggested that inference could be understood as a mental event whose proper function was to generate new acceptance states on the basis of old ones in a *reliably* correctness preserving way. I then suggested that we could view deductive inference as a case where the reliability in question was maximal: by preserving correctness at 'all' possibilities, deductive inference is as reliably correctness preserving as a transition between acceptance states could be.

Now that we have a characterization—indeed an analysis—of deductive inference on our hands, there are two important questions that need to be answered if we want to get a similar level of clarity about ampliative inferences.

> *Descriptive Question*: What is the structure of an ampliative inference, and in particular what are the structural relationships between ampliative and deductive inference?
>
> *Normative Question*: What does it take for an inference to be reliable in

> the ampliative case and, in particular, what would it take for an appreciated good ampliative inference to *be good*?

In the foregoing chapters I've tried to answer the analogous descriptive and normative questions for deductive inference. But here in the discussion of ampliative inference, I will focus *only* on the answer to the descriptive question. This is because I believe that most questions about the *nature* of ampliative inference in its relation to deductive inference can be resolved without embroiling ourselves in the complex and controversial question of what exactly it takes for such an ampliative inference to be good.

My account of the relation between deductive and ampliative inference springboards from two claims.

> Two Standards: Inferences can be subject to at least two distinct standards of goodness, and whether an inference is subject to a given standard depends only on the character of the inference (and not, e.g., on its contents).
>
> Mixed Inference: There are 'mixed' inferences in the sense that both of the standards of goodness of Two Standards apply to them to some extent.

I take Two Standards and Mixed Inference to be justified by simple examples. For example, consider two inferrers who make Inference 1 below, concluding with the truth of Goldbach's Conjecture.

> Inference 1
> The first even integer greater than 2 is the sum of two primes.
> The second even integer greater than 2 is the sum of two primes.
> ...
> The ith even integer greater than 2 is the sum of two primes.
> ---
> Every even integer greater than 2 is the sum of two primes.

Imagine pointing out to these inferrers that it is not obvious that the premises can *guarantee* the truth of the conclusion. For example, it is not yet clear from the premises why the $i + 1$st even integer *could not fail* to be the sum of two primes.

It seems possible for there to be an inferrer who regards this claim with indifference. Perhaps they are even rational to do so. This inferrer comes to

accept Goldbach's Conjecture on what appear to be enumerative inductive grounds. They might point out something like that the integers they consider are so many, or so representative, that it is perfectly reasonable for them to generalize from the finite cases to the infinite conclusion *even if* the truth of the conclusion is not guaranteed by the premises.

But we can also imagine an inferrer for whom the information that the premises do not obviously guarantee the truth of the conclusion *cannot* be regarded with indifference. That is, we can imagine an inferrer who mistakenly took themselves to have established the conclusion without any possible doubt (perhaps, e.g., they took for granted that there were only $\aleph$-many even integers greater than 2). As soon as a rational inferrer of this kind recognizes that their premises do not absolutely secure the truth of their conclusion, they would withhold from accepting it.

The difference between these two imagined inferrers cannot lie in the contents that figure as premises and conclusion of their respective inferences, since these are the same. And the difference *needn't* lie in their cognitive capacities to 'appreciate' an inference. Indeed, we can even imagine a single agent performing these two inferences at different times on the same day (perhaps having forgotten the first inference before they make the second). What seems to divide these inferrers is how the they 'connect' the premises to the conclusion, or how they 'view' the relationship between them.

Essentially the same idea has been noted by Boghossian, who claims it as a point in support of his Taking Condition.

> Intuitively [...] we are able to distinguish between a person who intends to be making a deductively valid inference versus someone who intends merely to be making an inductively valid one.
>
> A scientist need not be perturbed if we were to point out to him that some inference of his was not deductively valid, but merely inductively strong; but a mathematician would, and should, be perturbed.
>
> How, though, are we to capture the difference between the scientist and the mathematician, if not in terms of how they *take* their premises to be related to their respective conclusions?
>
> (Boghossian, 2014, 5)

Two Standards doesn't claim that the difference between the applicability of two standards of inference reduces to a difference in 'taking.' Rather

it merely emphasizes that there are at least two ways of inferring conclusions from premises, and that we should avoid running their distinct standards of goodness together.

An account of inference should illuminate where the standard of goodness applying to ampliative inference comes from, and how it relates to that for deductive inferences. I believe we get a clue for how to treat this relationship from Mixed Inference, which I claim to be supported by inferences like the following.[1]

Inference 2
r_1 is a black raven.
r_2 is a black raven.
…
r_n is a black raven.
q
$p \rightarrow q$

All ravens are black and p

I claim that it is possible for someone to (erroneously) infer the conclusion of Inference 2 from its premises here in a single step, and in a way that must constitute an ampliative inference in this sense: the inference is of the type which is a candidate for being good even if its premises don't necessitate its conclusion. The inference I am imagining is one that is performed by an agent who, for example, reacts with rational indifference upon being notified that it is metaphysically possible that their premises leave open that there are non-black ravens they have not yet encountered. That is, they react to this suggestion just the way one might rationally react to it when performing an enumerative inductive inference like the following.

Inference 3
r_1 is a black raven.
r_2 is a black raven.
…
r_n is a black raven.

All ravens are black

[1] Obviously this and subsequent inferences vastly oversimplify how enumerative induction works. I don't think this oversimplification will interfere with the utility of the examples to motivate the claims I am interested in.

So Inference 2 is subject to the kinds of standards that govern good ampliative inference. But I think it is also clear that the inference can be bad—indeed as I am imagining the inference being made, it *is* bad—precisely because it fails by the distinctive standards governing good *deductive* inference. That is to say, if someone pointed out to the reasoner I am imagining that there are metaphysically possible worlds where p is false even though q and $p \to q$ (perhaps alongside the claims about ravens) are true, they would recoil from their conclusion, and recoil from it *precisely as would* someone who had affirmed the consequent in a 'pure' instance of deductive reasoning.

Now one could claim that Inference 2 is in fact impossible in this sense: that one can't reason *directly* from the premises to the conclusion. One could insist that the reasoning would have to proceed in at least two steps, each of which would be either purely deductive or purely ampliative. I don't want to be completely dismissive of such a suggestion. For all I know, it may be true that as a contingent matter human reasoners have to break up reasoning steps in this way. But I regard the claim that it is *metaphysically necessary* for the inference to proceed in two steps as implausible. I don't see why an agent couldn't lump inductive and deductive moves into a single step in the way that human savants appear to lump together greater and greater chains of deductive reasoning into a single step. If this is right, although there are two kinds of standards that can be applied to inference, *some* inferences are to some extent governed by both standards, just as Mixed Inference claims.

I focus on the possibility mentioned in Mixed Inference because the intermingling of deductive and ampliative standards in a single inference points the way to a possible, and I think attractive, unification of all inference. On the resulting view, what unites inference is precisely that they are *all* subject to deductive and ampliative standards to some degree. It is just that in the familiar 'pure' cases of deductive or ampliative inference, the degree of the complementary standard is minimal or null.

There is in fact a natural way to develop this suggestion.[2] First, we get clearer on the structural conditions on good ampliative inference. Then, to accommodate mixed inference, we find a way to model these structural relations alongside those we have already uncovered for good deductive inference.

[2] Cf. Marcus (2020, 18,n.27), Marcus (2021, §5.5). The view to be developed here has natural kinships to those like the material theory of induction defended in Norton (2003, 2014, 2021)—though it is worth stressing that Norton's work primarily seeks to defend an answer to the Normative Question above, on which my view remains silent.

Lastly, having devised this model for mixed inference, we retrieve the structure of both pure deductive and pure ampliative inference as limiting cases.

To begin this process, let's look closer at the structural conditions on good ampliative inference. What distinguishes ampliative inference from deductive inference is that there can be circumstances at which the premises are true and the conclusion is false, but which nonetheless don't stand in the way of the inference counting as good. For example, the metaphysical possibility of the proposition that the $n + 1$st raven one sees is white even though all other ravens are black, may be irrelevant to some good inferences taking the form of Inference 3 above. Indeed, even if this proposition turns out to be true at the world where the inference is made, the ampliative inference may still have been perfectly good *qua* inference of its kind.

I will call a proposition *permissibly presupposed* if its negation is (metaphysically) compatible with an inference's premises, but incompatible with its conclusion, and is nonetheless rationally irrelevant to the goodness of that inference. For example, the proposition *that the* $n+1$*st raven I see is black* is among those permissibly presupposed in my hypothetical case of enumerative induction just given. This proposition has as a negation the proposition *that the* $n+1$*st raven I see is not black*. This proposition is compatible with the premises of the inference, but not with its conclusion, and it is nonetheless irrelevant to the goodness of the inference in question.

Now, we needn't specify which propositions *are* permissibly presupposed if we set aside the *Normative Question* above, as I propose to do. But to comfortably set that question aside we should also be cautiously flexible in allowing that permissibly presupposed propositions could be sensitive to a number of factors. These could include the context, the premises of the inference, probabilistic or other background information, the type of attitudes between which the inference mediates, and so on. Note that once we are flexible in these ways, the existence of *some* such set of permissible presuppositions for each individual ampliative inference is essentially guaranteed by the nature of ampliative inference. There are always propositions compatible with inferential premises whose possible truth could affect the goodness of the inference, and propositions whose possible truth could not affect it. Even if there is some vagueness here, any reasonable account of inference will have to accommodate the distinction, and so the existence of the latter set of propositions.

Now, in addition to the propositions that are permissibly presupposed in a

given inference, there are propositions that are *actually presupposed* in it. These are the propositions that a *reasoner treats* as permissibly presupposed in order for their inference to go through. To see that permissibly and actually presupposed propositions may come apart, consider how an inference of the following form could be made.

INFERENCE 4
r_1 is a black raven.

All ravens are black

If enumerative inductive inference is possible, we should expect there to be some suitable number of instances of black ravens n, perhaps in various suitable circumstances, the seeing of which would suffice to draw the general conclusion that all ravens are black. This means that there is some proposition of the form *it is note the case that: there are n seen black ravens, no seen non-black ravens, and at least one additional unseen non-black raven*, which is permissibly presupposed in an enumerative inductive inference like INFERENCE 3. But surely the proposition stating that *it is not the case that: there is* only one *seen black raven and several non-black ravens* is generally *not* permissibly presupposed in otherwise similar circumstances. Even so (and this is the point that matters for the present discussion), an agent may make INFERENCE 4 on this very presumption. That is, we can imagine a reasoner who makes INFERENCE 4 and who afterward encounters a second raven that is not black, but does not view this as in any way impugning the goodness of their earlier inference. They react to this outcome just would a lottery winner who initially judged their ticket wouldn't win on the basis of the incredible unlikelihood of this event: they take their earlier judgment to have been rationally based, the inference to be good, and so on.

Of course, the inferrer who performs INFERENCE 4 in this way is mistaken. The important point is that they are mistaken at least in part because they have *actually* presupposed propositions that are not *permissibly* presupposed. And just as there can be actually presupposed propositions that are not permissibly presupposed in an inference, there can be permissibly presupposed propositions that are not actually presupposed. This would occur if someone made a good inductive inference, but in such a way that if their attention were drawn to a permissibly presupposed proposition that their premises do not rule out, they would recoil from the inference's conclusion as if they had made a mistake (which, in *some* sense, they hadn't).

As already noted, when I talk of 'permissibly presupposed' propositions, I am passing the buck to a *normative* theory of good ampliative reasoning. Now when I talk of 'actually presupposed' propositions, I am *also* passing the buck, but this time to a *descriptive* account in the philosophy of mind about the nature of a cognitive relation underlying actual ampliative inferences. Actual presupposition is thus, for now, a schematic posit, much like the notion of 'taking' is for Boghossian. I do not take myself to be providing an account of this notion. I am only setting out a placeholder for a cognitive-relation with a particular role in inference that we can pick out more or less by ostension.

Without giving a full account of the relation of actual presupposition, I will rely on the claim that it has two key properties.

> Rationality of Presupposition: Actually presupposing propositions in the course of an inference is rationally epistemically evaluable. That is, one can be rationally faulted, in an inference, for actually presupposing propositions one is not permitted to presuppose.
>
> Presupposition as Passive Acceptance: Actually presupposing propositions in the context of an inference is functionally similar to an acceptance of those propositions as premises, up to the fact that presupposition is a passive relation.

Rationality of Presupposition should be straightforward, provided the cognitive relation even exists. The presupposing of propositions is posited to account for the kind of rational mistake present in a case like Inference 4 above where someone resiliently and unapologetically generalizes from too few cases. Again, *which* propositions it is irrational to presuppose and *why* it is not rationally permissible to presuppose them is something a normative account of good ampliative inference should tell us. But *that* there are propositions it is rationally impermissible to presuppose is a starting point for thinking about ampliative inference along the general lines I'm suggesting.

It is worth highlighting the hedge made by "*in the course of an inference*" in Rationality of Presupposition. Rationality of Presupposition is not assuming that the mental activity underlying actual presupposition is *always* epistemically evaluable. It may be that this relation occurs in other contexts. For example, perhaps what I am calling "actual presupposition" is a way of 'ignoring' possibilities that can occur outside the context of an inference alongside or as part of ordinary acceptance states. Perhaps one some-

times counts as actually presupposing various propositions while forming beliefs or suppositions as the result of inattention of cognitive impoverishment.[3] It may be that the question of the rationality of actual presupposition in these contexts cannot really arise (consider especially if it is possible to presuppose propositions in the course of engaging in idle suppositions). Even if this were so, that would be fine for my purposes. Rationality of Presupposition is neutral on these matters. It only assumes that in the course of an inference these forms of presupposition are epistemically evaluable. Hedged in this way, the claim seems straightforward, even if eventually *explaining* the claim may involve very hard work.

Presupposition as Passive Acceptance should be less obvious. But it can be bolstered by reflecting on what it is like to have one's attention drawn to a proposition during, or shortly after, one has made an ampliative inference in which that proposition is actually presupposed. I submit that *if* the question of whether a proposition is presupposable for the purposes of an inference arises, a rational agent is disposed *either* to accept the proposition (as one would a premise in the inference), *or* to abandon the inference. This should be relatively clear for inference between beliefs. Pointing out a proposition on whose truth the goodness of an inference turns, and which is unworthy of belief, must be grounds to defeat the rationality of believing the inference's conclusion—at least until other grounds are supplied. But it is true even of inference under counterfactual supposition: a proposition that was presupposed must be one that can be actively rationally supposed for the purposes of one's inquiry. If not, and if the permissibility of the inference hinged on presupposing that proposition, even the merely supposed conclusion is no longer supportable *as an inferential conclusion* in the context of suppositional inquiry.

Note that Presupposition as Passive Acceptance does not claim that presupposing propositions is a form of actual acceptance. On the contrary, the thesis states that presupposition is characteristically distinguished from an acceptance insofar as presupposition is a *passive* cognitive relation, unlike active acceptance. What I mean by passivity in this context is *representational* passivity or inactivity. In particular, I am assuming that one can presuppose a proposition in this sense without representing it—e.g., without having it occur to oneself—at all. The claim is that in spite of this, the rational role of presupposing propositions is like that of accepting premises in inferences. As

[3] Cf. the mental relation of presupposition discussed at Stalnaker (1984, 88).

I say, this is borne out by the reflection on how a rational agent relates to those propositions when they are called to mind: at that point the proposition is either accepted, or the rationality of the inference is recognizably defeated. Because of this parallel between presupposed and accepted propositions, so long as we are not concerned with the question of whether a cognitive relation is active or passive, we can provisionally treat the presupposing of propositions just as we would an acceptance of them.

With these two assumptions—Rationality of Presupposition and Presupposition as Passive Acceptance—we can give a reductive analysis of inference in terms of both crowding-out states and presupposition states that accounts for the possibility of mixed inference. In the process, we arrive at a general account of all inference. The analysis is the same as given for deductive inference in Chapter 5, with the exception that we allow the work originally done by active acceptance to be *shared* between active acceptance states and the passive acceptance relation of presupposition.

> A infers q from $p_1,\ldots,p_n$ if and only if there is some (perhaps empty) set of propositions S such that in A's cognition, states of crowding-out and presupposition are recruited for the purposes of an act of information extraction with the following character:
>
> (i′) A is sensitive to the specifying question of whether or not q while accepting $p_1,\ldots,p_n$ and presupposing the propositions in S,
>
> (ii′) in representing as per (i), A comes to crowd-out a representation not-q and $p_1,\ldots,p_n$ and S, and
>
> (iii′) A thereby comes to accept q alongside $p_1,\ldots,p_n$ while presupposing the propositions in S.

Several components of this analysis obviously need to be explained.

For example, what is it to be sensitive to the question of whether or not q while *both* accepting $p_1,\ldots,p_n$ *and presupposing the propositions in* S? While we may need a full account of the relation of presupposition to understand this sensitivity, we can bypass this with the help of Presupposition as Passive Acceptance. That thesis tells us that, as long as we are not concerned with the distinction between actively and passively accepting propositions, we can treat the passive relation of presupposing a proposition just as we would the active relation of accepting it. This means that, *as regards interaction with*

crowding-out states "accepting $p_1,\ldots,p_n$ and presupposing the propositions in S" is just like "accepting $p_1,\ldots,p_n$ and accepting the propositions in S".

So whatever presupposition is, as long as Presupposition as Passive Acceptance is true, (i′) and (ii′) will necessitate (iii′) in just the way I argued for deductive inference. If one comes to appreciate the impossibility of not-q alongside the premises *and* presupposed propositions, then the only way to maintain the acceptances and presuppositions while being sensitive to the question is to accept the conclusion.

There is still some important commentary to make on condition (ii′). It is no mystery how one could accept some propositions and presuppose others. It is accordingly also no mystery how one could be sensitive to a question while so-accepting and so-presupposing. But what is it to crowd-out not-q alongside some propositions *including* various presuppositions? The question may seem especially problematic because, as I've been emphasizing, presupposition is a passive relation which doesn't necessarily require anything like representing the contents of the presuppositions. Indeed, for all I've said, the set of presuppositions involved in a given inference could be infinite, and so not even in principle *representable*. But wouldn't crowding-out a set of propositions involving some presuppositions require representing them? That is to say, even if (i′) and (ii′) necessitate (iii′), why should we think that an ordinary reasoner could ever instantiate the condition in (ii′)?

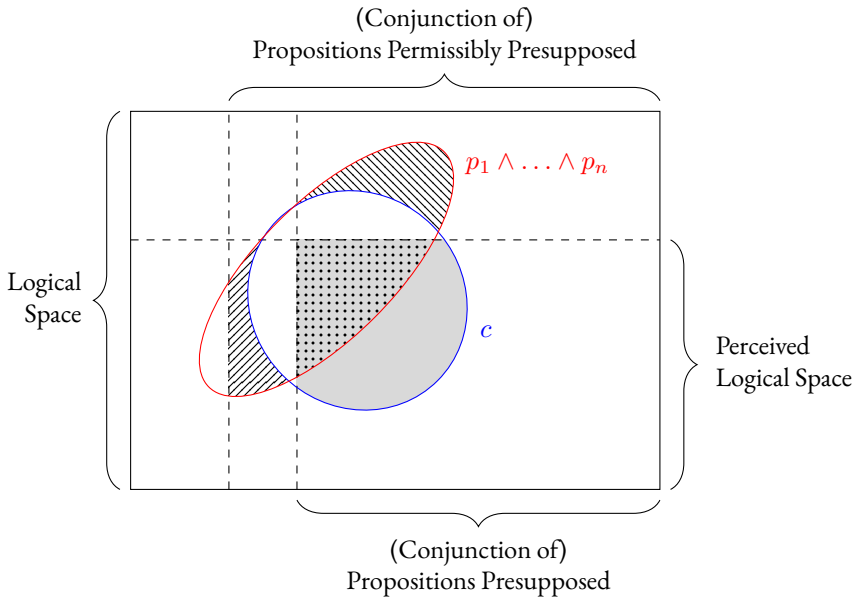
Here there is a happy confluence between the structures we've respectively imposed on crowding-out states and on presuppositional states. As I stressed in both Chapters 4 and 5, to crowd-out a proposition (or to crowd-out a collection of propositions) is not a representational state at all, let alone one in which the contents given by the propositions are represented. The relation of crowding-out is rather a representational mode on which the relationships between certain contents achieves a degree of clarity that begins to reflect the actual bounds of metaphysical modal space. In this context, crowding-out some propositional candidates for acceptance against presupposed contents is simply to represent in a way that reflects the actual bounds on modal space imposed by the presuppositions in question. This doesn't involve explicitly representing anything *about* the presupposed contents at all. So there is no special obstacle for thinking that the condition in (ii′) can hold for suitable contents, even for relatively unsophisticated reasoners.[4]

[4]In fact, it may well be that presupposition itself can at least sometimes be viewed as a

The basic idea behind the above analysis of inference is quite simple. The thought is that just as active acceptance of various claims as premises can help drive through acceptance of a conclusion in the reduced space for cognition created by crowding-out states, so too passive acceptance of claims can play the selfsame role. Passive and active acceptance interact on a par with the relation of crowding-out.

The structure of an inference supplied by the above analysis gives us the resources to see how there are two standards of goodness that can apply to an inference, and how these can occasionally intermingle significantly in a single inference. To illustrate, it may be helpful to see how the relevant structural relations could be modeled in a possible worlds framework for propositional content.[5] This can be seen in Figure 6.1.

FIGURE 6.1: Illustration of Two-fold Failure in Inference



component of a representational mode itself. Cf. the 'presuppositional' under-generating mode of representation of ink on a sheet given by Figure 4.5 of Chapter 4.

[5] The possible worlds framework has many limitations, and collapses some important distinctions needed to represent the full range of possible inferential transitions. But for the purposes of exhibiting the mere possibility of two-fold inferential failure it can be helpful.

We represent metaphysically possible worlds as points lying within the space of all such worlds, delimited here by the outer solid rectangle. Propositional contents in the worlds framework consist of the sets of metaphysically possible worlds where those contents are true. Accordingly a proposition (including a conjunction of propositions) is represented as a bounded subregion of logical space. In this case the conjunction of the premises of a given inference $p_1 \wedge \ldots \wedge p_n$ is bounded in red, and the conclusion c is bounded in blue. The inference takes place against backdrop crowding-out states of the reasoner, as well as against their backdrop states of actual presupposition during the inference. If a reasoner mistakenly crowds-out certain outcomes which are possible, this manifests itself as the agent's *perceived* logical space being narrower than it in fact is (represented by the region beneath the dashed horizontal line). When a reasoner actually presupposes a collection of propositions, this too delimits a sub-region of logical space (given by the region to the right of the rightmost vertical dashed line). Of course, actually presupposed propositions may not be *permissibly* presupposed in an inference. If a reasoner presupposes strictly more than is permissible, the space of worlds compatible with the reasoner's presuppositions will be a subregion of the space of worlds compatible with what is permissibly presupposed (this latter space being given by the region to the right of the leftmost vertical dashed line).

The overlap of an agent's presuppositions and their perceived logical space delimit what might be termed a space of *reasoner-relevant* worlds—in this instance given by the bottom-right rectangle. An inference can only take place if every reasoner-relevant world at which the premises are true (the dotted region) is also a world in which the conclusion is true (the shaded region), as holds in this instance.

But although the holding of these relations among reasoner-relevant worlds helps secure the possibility of an inference, and also encapsulates an important way in which an inferrer 'takes' their inference to be a good one, the *actual* goodness of the inference requires that premise worlds are conclusion worlds within a potentially broader space of *inference-relevant* worlds. These are metaphysically possible worlds compatible with the permissibly presupposed propositions: in this instance, the region to the right of the leftmost vertical dashed line.

In this example, there are two regions in this broader space of worlds where the premises of the inference are true and the conclusion is false, corresponding

to the two distinctive ways that an inference may fail to be good. The region hatched with north-west lines represents the worlds contributing to inferential failure in distinctively deductive terms. For example, supposing our figure is diagramming something like the mixed Inference 2, these could be worlds which reveal the reasoner to have engaged in the logical fallacy of affirming the consequent. In other words, these could be worlds where the premises including q and $p \rightarrow q$ of Inference 2 are true, but the conclusion p is not. The inferrer crowds out such worlds in the course of their inference even though they are metaphysically possible.[6]

The region hatched with north-east lines represents the worlds contributing to inferential failure in ampliative terms. These are worlds where the premises of the inference are true and the conclusion is false which were 'ignored' by being presupposed away. Note that not all worlds of this kind defeat the goodness of the inference: only those which are being presupposed away in an *illegitimate* manner. This is why the hatched region doesn't extend beyond the leftmost vertical dashed line. Again supposing we are representing something like the mixed Inference 2, the worlds in the hatched region here could be those with a certain number of black ravens and further non-black ravens, where such worlds cannot simply be presupposed to be non-actual in the context of the inference, but must be actively ruled out by premise-taking acceptance states.

In a good inference, neither of the hatched regions would exist. First, either the permissibly presupposed propositions would line up with the presupposed propositions (so there is no 'room' for ampliative failure) or, if they do not line up, the goodness of the inference would not depend on the region where they fail to overlap (i.e., the region between permissibly presupposed and presupposed contains no world where premises and true and conclusion false). Second, either perceived logical space lines up with true logical space (so

[6] Incidentally, though there are no such worlds in this diagram, worlds in the top-middle rectangle where premises are true and the conclusion is false would *also* contribute to distinctively deductive failure. These are in some sense also ruled out inappropriately by presupposed propositions, but I take a reasoner's perception of logical space to be an overriding means of ruling out worlds in these cases, and so to take precedence. That is, were the agent not to have presupposed what they did, these worlds would have clearly still been ruled out for the reasoner, whereas if the reasoner 'perceived' more of true logical space it is an open question whether their presuppositions would have been maintained. That is my impression, at least. I suppose if that impression is wrong there may be a class of worlds that contribute to deductive and ampliative failure at the same time. I am open to this possibility if it really arises.

there is no 'room' for deductive failure), or the goodness of the inference does not depend on their lining up (i.e., the region of logical space outside perceived logical space contains no world where the premises are true and the conclusion is false).

As promised, not only have we accounted for the possibility of mixed inferences but inferences quite generally, with 'pure' ampliative and deductive inferences treated as limiting cases. Pure deductive inference is simply inference where no propositions are actually presupposed (or perhaps: any propositions actually presupposed have no bearing on the goodness of the inference). And pure ampliative inference is simply inference where the deductive component of inference is as trivial as possible, and the most important work in securing the inference is effected by presupposed propositions. This last idea can get distorted a little in the worlds framework, were we lack the ability to represent fine-grained distinctions among the ways an agent may crowd-out various impossible propositions, or not. And even abandoning the worlds framework wouldn't obviously lead to a precise characterization, since we don't have a 'metric' of simplicity over impossible propositions or collections of propositions. Nor is it clear such a general metric could be devised, given the psychological variability of deductive appreciability.

In spite of all this, the rough idea is hopefully clear: just as I could ampliatively infer some conclusion q from premises $p_1,\ldots,p_n$, I could also in-principle infer any appreciable deductive consequence q' of q from the *same* premises $p_1,\ldots,p_n$ in a single step. But in general, the harder to appreciate the consequence q' in question—that is, the harder it would have been to deductively get from q to q' in a single step—the harder it will be to infer q' from $p_1,\ldots,p_n$ in a single step as well. The most 'pure' ampliative inferences are the ones where the 'deductive distance' from the premises $p_1,\ldots,p_n$ to the conclusion is minimized. For example, it may be no more complex than something like Modus Ponens. Or maybe, in the truly limiting case, no more complex than an inference from a proposition to itself. But minimized though it may be, there is a sense in which a deductive component is *always* present in any inference. Even in 'pure' ampliative inference, crowding-out relations are always present, and driving the acceptance of a conclusion on the basis of various premises.[7]

[7] This gives my final answer to a worry of MacFarlane (among others) discussed in Chapter 3 that deductive inference is abnormal and takes training to perform at all. On my view, he could

One might be tempted to say (perhaps as an objection): this is a view on which all inferences are *really* deductive, precisely because the mechanisms of deduction are ubiquitous. Strictly speaking this charge is inaccurate. There is a firm distinction in the view between two very different standards of goodness, and those differing standards apply to intuitively ampliative and intuitively deductive arguments respectively in just the measure one would expect them to apply. That is, bad ampliative inferences, bad deductive inferences, and even bad mixed inferences, are each criticizable on precisely the diverging grounds we would expect. The standards that apply are different in kind, and apply to the inferences on account of the presence of significantly structures belonging to the inferences in question.

There is, however, a grain of truth to the charge. The grain of truth is that the view just given places ampliative and deductive inference at a very slight remove from each other, in that presupposed and accepted premises in an inference are essentially interchangeable (up to worries about cognitive load, for example). In any ampliative inference, the work done by presupposition could (bracketing worries about finitude) potentially be shifted into an acceptance state to render the whole process deductive. The distinction between deductive and ampliative inference boils down to a distinction between actively accepted premises and passively ignored possibilities.

Though this version of the charge would be correct, I cannot see how it could constitute an objection to the view, since ampliative and deductive inference are *in fact* closely related in just the way the view posits. To give one important class of examples, *any* ampliative inference mediating between belief states could have equally well been a deductive one: instead of inferring a conclusion from some premises, one could rationally come to believe the conditional linking the conjunction of the premises and the conclusion. (One way to do this would be by supposing the premises, ampliatively inferring the conclusion under supposition, and on the basis of this process coming to believe the conditional. But perhaps there are other routes as well.) Then one could deductively infer the conclusion with the help of the new premise. It

be right in at least this sense: it may be extremely unusual for an ordinary agent to infer without presuppositions, and it may take instruction to learn to dispense with those presuppositions while engaging in reasoning (e.g., when one learns to do proofs in mathematics or logic for the first time). But even so, deductive processes effectively *underlie* all inferential transitions. So a study of deductive inferential goodness is not the study of an exotic and specialized mental event, but an essential *component* of an equally essential process of reasoning.

seems that this overall process of coming to deductively accept the conclusion through an explicit premise plays an equivalent rational role to the acceptance through ampliative inference without the help of the linking conditional. If it was rational to presuppose the truth of any propositions which actually secured the truth of the consequent against the backdrop of the other explicitly accepted premises, then the conditional is rational to explicitly believe as well. This is, in a way, a variation on the point I made above about how passively accepted propositions play a rational role in inference equivalent to actively accepted ones.

So I see the links created between ampliative and deductive inference as a feature, not a bug. This, of course, is hardly the end of a defense of the account of inference on offer, but just the beginning. With the account in hand, the real defense will come by turning back to the question that motivated this chapter: can the lessons about deductive inference defended in Chapter 5 be extended to all inference on this account, at least where those extensions appear appropriate?

6.2 Lessons Extended, Loose Ends

Deviant Chains, *Moorean Incoherence*, *Sophistication*, and *Rationalizing without Regress* seem to apply to ampliative inference just as much as they do to deductive inference. Let's begin by showing how the account of §6.1 respects these connections.

(I) *Deviant Chains: An account of inference should illuminate why mere causation among acceptance states is insufficient for inference to take place, and illuminate what connection between acceptance states* is *required for it.*

The extended account of inference preserves the virtues of the account of deduction: in any inference, the very thing which distinguishes it from a mere chain of acceptances—namely the situated presence of crowding-out relations—is also the thing which constitutes the acceptance of a conclusion 'in the right way.' Since there is no causation appealed to in the reduction of even ampliative inference, there are no worries about causation at all, let alone any about deviant causation.

(II) *Moorean Incoherence: An account of inference should explain the IMP—*

that is, how a rational tension necessarily arises between any conscious inference and the conscious judgment that the inference is a bad one.

Recall that the IMP is explained in *deductive* inference by two claims: first, by the claim that in the context of an inference states of crowding-out are epistemically evaluable; second, by Rational Sensitivity to Assessment.

> Rational Sensitivity to Assessment: In a rationally coherent mind, a conscious epistemically evaluable process or state that is rationally judged epistemically deficient will in typical conditions yield to the judgment.

Jointly, these entail that one must be guilty of irrationality if one judges of a consciously performed inference that it is a bad one in typical conditions.

An account with exactly the same form can be given for inference generally granting Rationality of Presupposition. This assumption tells us that presuppositional states, just like crowding-out states, are subject to epistemic evaluation in the context of an inference. Accordingly, to judge of an inference that it is bad must involve judging either that the presuppositions in the inference are impermissible, or that the representational modes involved are inadequate, or both. No matter how one judges, either one judges irrationally, or by Rational Sensitivity to Assessment the presuppositions or representational modes that help constitute the inference cannot rationally be maintained.

Note that unlike with crowding-out states, we don't yet have an explanation of exactly *why* relations of presupposition are evaluable in this context. But this fuller explanation will likely have to wait for a complete answer to the *Normative Question* of what makes for a good ampliative inference. Only such an account can fill in precisely what purposes are being served by the relation of presupposition in securing a *reliable* extraction of information from information-bearing states. So, allowing that we cannot be expected to answer the *Normative Question* here, we have as detailed an explanation of the IMP as we could hope for.

(IV) *Sophistication: An account of inference should explain how inferring is possible for relatively unsophisticated reasoners like young children, or explain how 'inference-like' activities of such reasoners do not properly count as inferences.*

The only difference between the account of deduction and general inference is the integration of presupposition. Presupposition is, as I have been stressing, a characteristically *passive* cognitive relation. So although presupposition is in many ways unlike the relation of crowding-out, it shares with it the crucial feature of *not* being a representational state. This was the key feature that resolved worries of sophistication for deductive inference. So exactly the same account can be given for the ampliative case as well. Ampliative inference requires *neither* that the inferrer represent an explicit connection between premises, presupposition, and a conclusion, *nor* presuppositions themselves. It only requires that one 'ignores' certain possibilities while representing the premises and conclusion with certain representational modes. In other words, it continues to be true that no conceptual sophistication is required to perform an inference beyond that needed to entertain the premises and conclusion.

(V) *Rationalizing without Regress: An account of inference should explain the rationalizing feature that distinguishes inference from a mere succession of attitudes. And it must explain how it rationalizes acceptance states, sometimes compelling them, without creating a problematic form of regress (framed in slightly different ways by Carroll, Boghossian, and Kripke).*

Again, because in passive presupposition one *need not represent*, it must of course be possible to be rational in presupposing without inferring the presupposition as the conclusion of an inference, or basing it explicitly on a belief, and so on. Note, of course, this is merely to say that presupposition has the right *form* to avoid regress. Saying exactly how presuppositions can and do become justified may be a complicated matter, and one that will have to await a suitable answer to the *Normative Question*.

Avoiding regress in this way avoids the problem for constitutivist views raised by Blake-Turner (ms./2021) that I alluded to at the outset of

this chapter. The objection Blake-Turner raises for Valaris is that the latter's account cannot explain how regress is avoided in ampliative inference. For Valaris's account says that in inference, one judges that premises provide *conclusive support* for their inferred consequence. And it is recognition of this conclusive support that allows that recognition to plausibly help *constitute* the judgment of the conclusion. As Blake-Turner notes, this seems to leave ampliative inference unaccounted for. And it is not clear we can expand the constitutivist proposal by weakening the recognized support relation:

> If some weaker relation of support is allowed, then it is false tha[t] an agent's believing both the premises and corresponding taking state of an inference constitute her believing its conclusion. I can, rationally and attentively, believe both:
>
> (6) There has been snow on the ground every January 1st in Niseko for the last 50 years.
>
> (7) (6) supports "There will be snow on the ground next January 1st in Niseko."
>
> Without believing:
>
> (8) There will be snow on the ground next January 1st in Niseko.
>
> But I can infer (8) inductively from (6).
>
> Blake-Turner (ms./2021)

One thing Blake-Turner is pointing out here is that in rational ampliative inference there appears to be a kind of rational discretion in whether to accept a conclusion. He bolsters this claim with a discussion of how pragmatic or moral encroachment, which 'raise the stakes' of accepting a consequent, can make it rational to withhold from drawing the consequent as a conclusion.

The present view can account for all these facts. First, on the present account, there is *always* a necessitation relation present in inference, including ampliative inference. It is just that in ampliative inference this relation does not hold between premises and a conclusion, but between premises *supplemented* by presuppositions and a conclusion.

> Note that Valaris cannot obviously take this route precisely because he is committed to taking-relations being representational states (unlike with crowding-out states). In this instance, this would seemingly require reasoners to explicitly represent every presupposed proposition of every inference, which is highly implausible.
>
> The present view can also explain why ampliative inference has a characteristic discretionary character that deductive inference may lack. This is precisely because the discretionary character of such inferences derives directly from the rationally discretionary character of presupposition. Presuppositions are *rationally permitted* or not, but not obviously always *rationally required*. Accordingly, even if one is permitted to presuppose some proposition (e.g. one that undergirds the support relation reported in Blake-Turner's (7) above), one may equally well be rationally permitted in some circumstances (perhaps those with raised stakes among them) to be cautious and not presuppose that proposition when considering an inference. Accordingly, one may explicitly judge the premises to inductively support a conclusion, and accept the premises, without drawing the conclusion. One may even 'appreciate,' 'take,' etc. the premises to inductively support a conclusion without drawing it, in the sense that one may see the connections between premises and conclusion given presuppositions of certain sorts, without yet having taken on board the presuppositions. As long as one has rationally refrained from making those rationally discretionary presuppositions yet, one will have refrained from drawing the conclusion, and done so in a rationally permissible way.

This shows that the account of inference I've offered neatly extends the accounts of conditions on deductive inference that ampliative inference clearly shares. It explains *Deviant Chains*, *Moorean Incoherence*, *Sophistication*, and *Rationalizing without Regress*, and in essentially the same ways for the deductive and ampliative settings. But the account can also illuminate why the remaining conditions *Insufficiency of Knowledge* and *Small Steps* do not transpose as clearly to inferences generally. Consider the first condition.

(III) *Insufficiency of Knowledge: An account of inference should explain why knowledge that an inference is good is generally insufficient to enable one to rationally perform the inference.*

It is clear that knowledge of the goodness of a deductive inference does not typically position one to perform it. But can knowledge of the goodness of an ampliative inference do so? Here I find things less clear.

Suppose I have seen enough black ravens to inductively infer all ravens are black, but I refrain from drawing that conclusion. I'm simply unsure I've gathered enough evidence. But now suppose I am told by a credible authority, who I fully trust, that I have seen a safe number of ravens to justify the conclusion that all ravens are black. Can I now judge that all ravens are black in the right way, *merely* on the basis of the ravens I have seen? It's certainly clear I can now rationally get to the conclusion that all ravens are black. But can I get there by performing the original ampliative inference I was refraining from making, or only (in normal cases) by performing a new inference, with an added premise given to me through testimony? I submit this is hard to evaluate.

My account cannot resolve this issue. But it should not be expected to. This is because I've passed the buck on accounting for the nature of actual presupposition, and an account of that mental relation is needed to resolve the question of whether active acceptance of a proposition is enough to enable someone to (actually) rationally presuppose it.

What an account like mine should do, however, is explain the felt difference between the deductive and ampliative cases vis-à-vis *Insufficiency of Knowledge*: Why is it *clear* in the deductive case that knowledge is insufficient, and harder to tell in the ampliative case whether it helps? And the account can explain this. The difference between an 'original' ampliative inference and a 'new' one with added premises in cases like the above boil down to a difference between presupposition and acceptance. And the difference between passive acceptance (presupposition) and active acceptance is, as we've had occasion to note, relatively slight. Indeed, up to the question of representational passivity and activity, these two kinds of states appear to have similar rational roles. By contrast, the distinction between an active acceptance and a representational mode is vast. A representational mode is not, nor could it ever be confused with, a representational state.

Because of this it should be expected that we would have less clear judgments about the applicability of a principle like *Insufficiency of Knowledge* to ampliative inference. A fuller account would take a stand on this application. But that fuller explanation would have to say much more about presupposition. Because my account is largely neutral on the specification of presuppo-

sition relations, it should be combinable with any reasonable elaboration of them. And that suffices for my purposes here.

Consider next condition (VI).

(VI) *Small Steps: An account of inference should help explain when, and why, single-step inferences are rationally available for a given inferrer—and in particular why deductive inferences tend to proceed in relatively small such steps for ordinary human reasoners.*

This condition makes good sense in a focus deductive inference, since we have a rough but intuitive sense of what a 'small deductive step' comes to. But what makes an ampliative step a greater or lesser one? Here I think the applicability of a notion of 'size' to an inference is significantly less clear. Of course, one thing we can say is that some ampliative inferences are more of a 'stretch'—for example reasoning to all ravens being black from seeing m black ravens may be 'safer' than doing so from seeing n black ravens when $n \ll m$. But this kind of 'ampliative distance' in enumerative induction is clearly different in kind from that of deductive distance: it is not like more ambitious ampliative inferences are 'harder to see' in the *same way* that a complex Ramanujan-style inference is.

Schechter (2019, §3) argues that there is still a parallel between deductive and ampliative inference here, giving as an example that it would be irrational for ordinary reasoners like ourselves to infer the truth of a scientific theory from a large collection of experimental data directly and in a single step. While I think this latter claim is true, I also think its implications for the nature of inference are far from clear. If we are imagining a typical chain of reasoning that leads from data to overarching scientific theory condensed into a single step, what would typically be condensed is a long string of *both* ampliative and deductive inferences. That is, the hypothetical single-step inference would have to be a *mixed* inference in the sense I used above. This makes it harder to see if what is really 'hard' about the case distinctively concerns ampliative inference.

Another important feature of Schechter's example is that the premise set is huge and unwieldy (and the conclusion may be as well, depending on the theory). It is not only hard to imagine someone inferring from the premises in this case. It is hard to imagine them simultaneously and consciously believing those premises. This raises the concern that some of the difficulty sensed in the example is not tracing to a difficulty in *inferring* proper, but rather a difficulty in adequately forming an acceptance state that is a prerequisite for inference.

I suspect there may be variants on Schechter's case that factor out the foregoing problems to some extent, and where some sense to the difficulty of a distinctively ampliative inference remains. But while I agree with Schechter that there are probably hard ampliative inferences that need some accounting for, I disagree with his eventual claim that we should look for a *unified* explanation of these phenomena,[8] as there seems to me to be important contrasts between them. A key difference is one just discussed in *Insufficiency of Knowledge*. There are some pure deductive inferences, with small numbers of easily graspable premises, where it can be *extremely hard* to perform the inference in a single step, while it is no special challenge whatsoever to believe an entailment holds between the inference's premises and its conclusion (for example on the basis of testimony). It is not clear that we have parallel examples of this in the ampliative case. Granted, a variant of Schechter's case with deduction factored out may involve a single-step ampliative inference that is hard to rationally perform. But it seems to me that the typical cases of this form will be ones in which it is commensurately challenging to rationally come to accept that the ampliative entailment relation holds (as seems to hold in Schechter's case because of the sheer number of premises involved).

So while the *extent* of the differences between the 'step-sizes' of ampliative inference and deductive inference is certainly debatable, what seems clearer to me is that there are important qualitative contrasts between them. And the theory I have put on offer has the flexibility to respect these differences, and even predict them. On my view, the ease or difficulty in performing a 'pure' deductive inference typically traces to the ease or difficulty of representing while crowding-out certain propositions. And we have experimental evidence that it is *highly* cognitively demanding to correctly crowd-out complex relationships among contents even when it can be perfectly easy to rationally believe the relevant propositions are impossible. The ease or difficulty in performing a 'pure' ampliative inference, by contrast, would instead trace to the ease or difficulty in *rationally actually presupposing* various propositions. And this is simply a different kind of cognitive relation.

This difference makes room for the possibility of different step-sizes in ampliative reasoning, without identifying them with the step-sizes involved in deductive reasoning. It makes room for degrees of ampliative difficulty because, while it is probably not especially cognitively challenging to presuppose propo-

[8] Schechter (2019, 158–9).

sitions of virtually any complexity, it can arguably be harder or easier to *rationally* presuppose them. But this difficulty is different from that involved in performing deductive inference.

In sum: while there may be greater or lesser steps in both inductive and deductive inference, these steps seem different in nature, and could respond differentially to things like rational testimony. The account I've given is well-positioned to explain all of these facts.

I think all the foregoing work of this section shows that the reduction of inference I've put on offer at least has the *structural* features needed to extend the account of deduction to one for ampliative inference in a satisfying way. Still, the account is one which still awaits a fuller account of states of presupposition, as well as an answer to the critical *Normative Question* to be complete and fully assessable. By way of conclusion, I want to say a few more things about why this supplementation is needed, what shape it must take, and why I am not providing it here.

The key to extending an account of deduction to ampliative inference, I have claimed, is an account of a passive mental relation of presupposition satisfying two constraints: Rationality of Presupposition and Presupposition as Passive Acceptance. Collectively, these principles attribute key features of crowding-out states and accepted premises to presuppositions. Presupposition as Passive Acceptance allows presuppositions to drive through inferences like accepted premises, but also allows the relation of presupposition to be passive and non-representational like crowding-out states. And Rationality of Presupposition allows presuppositions to be distinctively epistemically evaluable in the context of an inference, again without being explicitly representational, just as with crowding-out states. By allowing presuppositions to occupy some of the roles of accepted premises, and some of the roles of crowding-out states, we can slot them into my constitutivist analysis of deduction without disrupting its key virtues.

While I think it is intuitive that there is *some* cognitive relation involved in inference which satisfies Rationality of Presupposition and Presupposition as Passive Acceptance, these theses are not so intuitive that we can rest content without seeing a developed account of the cognitive presupposition relation. There are two things in particular that I think need further explanation, corresponding to each of the two theses I've leaned on. Concerning Rationality of Presupposition, it is worth highlighting this partic-

ularity: this thesis labels presupposition states as irrational *for the purposes of inferring*. This is essential as one can rationally suppose, and so probably rationally presuppose, whatever one wants in counterfactual supposition. The rationality of presupposition is tied to its role in inference as facilitating reliable or safe information extraction. But these ties to reliability or safety, to be fully understood, require an answer to the *Normative Question*. We cannot know what the reliability in question is until we know exactly which ampliative inferences count as good and which do not. (Compare: showing crowding-out states could be rationally evaluable in the context of an inference *did* require an understanding of exactly to what extent they contributed to the extraction of information.) As I've been careful to flag, though, answering the *Normative Question* is simply not possible in the scope of this book.[9]

As regards Presupposition as Passive Acceptance, the particularity we need to be mindful of is a tension between the acceptance-like role of inference and its passive character. Can we spell out what it is for a mental state to 'ignore' various possibilities or circumstances without this collapsing into an 'active' form of belief or other form of acceptance that requires explicit representation of contents? This will depend *both* on our account of presupposition *and* our account of belief or acceptance more generally. It should be clear that broaching these topics would take us far afield from our present investigation. But the success of the present account is tied in some respects to whether the investigation of these further topics pans out.

Given these uncertainties, what can we take ourselves to have established? My focus in this book has been on deductive inference, since it is this mental activity that I believe we profit from seeing deductive logic as investigating. But we cannot claim to have a firm grip on deductive inference until we at least have *some* sense of how it relates to inference more broadly. This is because the unusual features of deductive inference—indeed the very unusual features which I will, in Part II, *apply* to logical problems—are sometimes clearly shared, but sometimes not clearly shared, with inference more broadly. Still, resolving all

[9] We also need to make sure that when we specify what makes presupposition rational in the inferential context, we do so without mentioning inference, lest we give up the reductive ambitions of the view. In the deductive case, the corresponding problem was avoided by adverting to the recruitment of crowding-out relations in an activity of maximally safe information extraction. So the issue here is likely to boil down to the question of whether a laxer notion of 'safety' can be spelled out in non-inferential terms. Thanks to Chris-Blake Turner for alerting me to this point.

questions about inference would balloon our investigation to unwieldy proportions.

Accordingly what we need, and what I hope to have provided, is some indication of how a fuller investigation into ampliative inference could be undertaken, and some reasons for thinking that the investigation will yield results consistent with the account of deductive inference I've put on offer. I think that while the account of this chapter passes the buck at key junctures, it still does at least that much. It shows how a mental relation with an intuitive basis can be adjoined to the account of deduction to yield a general analysis of inference with just the features we want. The account gives neat, parallel accounts of the features shared by deductive and ampliative inference, and also gives the tools to explain the differences between them when they arise. In the fullness of time, we will want to know that the intuitive relation has a firmer basis in both our descriptive accounts of it in the philosophy of mind, and our normative accounts of relations of rational support for ampliative inference. But for present purposes, we have made a good enough start to lean on the account of deduction of Chapter 5, and explore the implications of accepting it for the foundations of logic. It's to this task that I now turn.

Part II

Applications

CHAPTER 7

First-Order Validity & A Reduction of Consequence

In Chapter 2, I presented a skeletal account of deductive inference as a mental process whose proper function is to appreciably generate new acceptance states on the basis of old ones in a maximally reliable, truth-preserving way. This account left two key aspects of deductive inference unspecified. First, the account did not specify the kinds of worlds over which a deductive inference must preserve truth. Second, it did not specify what it meant for an inferential transition to be 'appreciated.'

Drawing on crowding-out relations from Chapter 4, I gave a reductive analysis of deductive inference in Chapter 5 that fills in these two gaps. Crowding-out relations give the sense in which an inference is appreciated. This relation reduces the space of possible thoughts for an agent, sometimes thereby settling an agent's deliberation on a question. And since *good* inference settles deliberation through *correct*, or representationally adequate, crowding-out relations, deductive inference must aim at a transition preserving truth at metaphysical possibilities. This is because all and only such transitions are capable of being the product of representationally adequate crowdings-out.

As we close outstanding questions about our skeletal conception of inference, however, we raise new questions about the associated skeletal conception of logic. I said in Chapter 2 that one central conception of logic should be the linguistically mediated study of good deductive inference. It now turns out that the feature of good deductive inference that logic studies is metaphysically necessary truth-preservation. But this raises doubts that logic in my sense could possibly accord with logical practice, especially when considering the standard Tarskian model-theoretic machinery used to characterize first-order

validity. Familiarly, although that machinery tracks truth-preservation across a range of 'cases,' those cases are not naturally construed as metaphysical possibilities. So how could first-order model-theoretic consequence be a form of logical consequence, in *my* sense of logic?

I start to address this concern in §7.1. I begin there by characterizing a class of linguistic properties under the heading of 'modalized first-order form.' I then give an informal argument that first-order model-theoretic consequence tracks transitions among contents expressible by first-order sentences that preserve truth at all metaphysical possibilities in virtue of bearing that form. This shows that model-theoretic validity tracks one 'true' form of logical validity on my view. The result reveals that classical logic is the distinctive logic of certain 'semantically well-behaved' discourses, of which mathematics provides a central case, though perhaps the only clear one. This result is meant both to do justice to the importance of first-order model-theoretic techniques, while also giving a clearer understanding of the limitations of those techniques.

Once this is accomplished, we can actually see that my proposed construal of logic does more than merely accord with logical practice. Using Etchemendy (1990, 2008) as a foil, I argue in §7.2 that by characterizing logic in terms of deductive inference while providing a reduction of deductive inference, we have provided a reduction of, and framing for, logical consequence relations that holds out the promise of fruitfully reducing logical questions to non-logical ones. For example, the analysis and its setting within a broader framework of inference gives us the resources to explore non-question-begging answers to questions about the validity of controversial inference rules. The account does this by transforming disputes about such rules into (often challenging) non-logical questions in the domains of philosophy of language, philosophy of mind, linguistics, and metaphysics. Near the end of §7.2, I briefly note how this transformation takes effect for disputes over Excluded Middle and Ex Falso, which are preparatory for more detailed case studies in Chapters 8–11.

7.1 First-Order Model-Theoretic Consequence

The sentences of a first-order language $\mathcal{L}$ are built recursively from variables, constant/function symbols, predicate symbols, truth-functional connectives, quantifiers, and (optionally) an identity sign, in a familiar way. A (first-order)

model $\mathcal{M}$ of a first-order language $\mathcal{L}$ consists of a non-empty, set-sized domain D of objects, and extensions built from elements in D for each constant, function, and predicate symbol (besides identity) in $\mathcal{L}$. Truth of a sentence of $\mathcal{L}$ in a model is defined in the familiar way via an inductively defined satisfaction relation that reflects classical assumptions about the behavior of connectives and quantifiers. A sentence ϕ is a logical consequence of a set of sentences Γ ($\Gamma \models \phi$) just in case any model rendering Γ true also renders ϕ true. A sentence ϕ is logically valid ($\models \phi$) just in case ϕ is true in all models. In what follows, I'll typically focus on first-order validity for simplicity, though the relevance to first-order consequence will be clear.

Truth-in-a-model is a mathematically defined relation between a set-theoretic abstraction—a model—and a sentence-type in an artificial first-order language. What can the classification of first-order sentences on the basis of this mathematically defined relation have to teach us about *true validity*? According to the view I've been defending, true validity is assessed relative to some type L of linguistic properties (where I will continue to leave open whether or not it is only a privileged subset of linguistic properties that should count as 'properly logical'). A sentence is L-valid just in case it expresses a necessary truth in virtue of the L-type properties it possesses.[1]

Accordingly, the question we would like to ask goes something like this.

(Q) Is there some type of linguistic property L such that, for all interpreted first-order sentences ϕ:

$$\phi \text{ expresses a metaphysical necessity in virtue of its } L\text{-properties} \Leftrightarrow \models \phi\,?$$

We must ask our question about *interpreted* sentences (where I so far use the term "interpretation" informally) because only sentences with adequately specified semantic properties could ever express the kinds of truth-conditional contents that could figure as the beginnings or ends of good inference, and that can be assessed for necessary truth or truth-preservation in connection with such inferences.

[1] I mean for this formulation to allow that the *particular* properties in virtue of which sentences expresses necessities my vary from sentence to sentence. The 'type' of property L may be thought of as determining the particular properties given a sentence. We'll see an example of this very shortly.

The problem is that so far (Q) is meaningless or has a vacuous answer if the 'interpretations' in question are provided by first-order models. These models only provide expressions with extension-level semantic properties like referents, predicate-extensions, and truth-values. A single model provides semantic information about at most one world—presumably, actuality. But to ask questions about true validity, we need to ask about the expression of metaphysical necessities. And to do that, we need to ask questions about the kinds of truth-evaluable objects of speech and attitudes that could figure as the starting and endpoints of genuine inference. So we need information about *truth-conditions*—truth-values relative to metaphysically possible worlds. Even if first-order sentences are regimented from natural language sentences that have such truth-conditions, the regimentation involves processes of abstraction and idealization that can in-principle distort semantic properties of a natural-language source.[2] Most importantly, that process abstracts from the modal properties of natural language sentences' assertoric contents that bear most directly on logical matters. So to even properly formulate a question about 'true validity' in the sense I've offered, we have to undo some of this process of abstraction.[3]

While there are several paths to take in response to this issue, the simplest is to consider 'interpretations' that simply generalize to the world-relative setting the techniques of model-theoretic assignment of semantic properties.

> $\mathcal{I}$ is a *modalized interpretation* of a first-order language $\mathcal{L}$ just in case $\mathcal{I}$ is a function from metaphysically possible worlds w to models $\mathcal{M}_w$, such that the domain of $\mathcal{M}_w$ is drawn from objects existing at w.[4, 5]

We know that to adequately extend ordinary model-theoretic interpretation to the truth-conditional setting, we need interpretations that both behave like

[2] See Glanzberg (2015) for a helpful discussion.

[3] Essentially this point is fully appreciated, and helpfully discussed, in Gómez-Torrente (2008).

[4] There could be concerns about merely contingent objects figuring in the domains of any models that are the values of an *actual* modalized interpretation at non-actual worlds. In this case, I will assume that ersatz entities stand in the domain of the model to fulfill the relevant function.

[5] Note that the domain of a modalized interpretation is *not* variable—it is fixed by the space of metaphysical modality. This contrasts with the role of worlds in, say, Kripke models (see Chapter 8).

a first-order model at the actual world and also assign truth-conditions relative to non-actual worlds. Modalized interpretations satisfy these conditions in the simplest way possible, by behaving like some first-order model at every possible world, using some set of that world's objects as a domain.

Though modalized interpretations are a natural, simple class of interpretations to consider, it is worth noting that one may wish to consider narrower classes of interpretations. The above characterization made no assumptions about the relationships among the models that capture the modalized interpretation's behavior relative to various possibilities. But one might have an interest in establishing such relationships. As regards quantifier interpretation, for example, nothing I've said yet requires modalized interpretations to give quantifiers constant domains from world to world, even if the range of objects existing at various worlds does not change. Nor have I assumed that constant symbols receive a constant or 'rigid' denotation across metaphysical possibilities. And so on.

I will say more about the importance of some of these choices soon. For now, modalized interpretations supply one way of giving truth-conditional content to first-order sentences broadly in line with model-theoretic interpretation. So they give us one class of 'interpreted' sentences that could give significance to our earlier question about the relationship between model-theoretic validity and a form of true validity. We can accordingly reformulate it as follows:

(Q′) Is there some type of linguistic property L such that, for all first-order sentences ϕ *given a modalized interpretation*:

$$\phi \text{ expresses a metaphysical necessity in virtue of its } L\text{-properties} \Leftrightarrow \models \phi\,?$$

This is now a sensible question, and the answer to *it* is "yes". The property type L in question can be identified by extracting linguistic commonalities from the class of modalized interpretations. But I will need to say quite a bit more about how I am thinking about linguistic properties in general before specifying the property type that is relevant to the above equivalence.

I will focus here on *syntactic* and *base semantic* properties of sentences. This immediately excludes, among other things, any pragmatic properties these sentences might have (considered here as properties of linguistic usage which are not relevant to the determination of truth-conditions).

Syntactic properties of a sentence are properties concerning orthography, and grammatical type and composition. These include properties like that of being a sentence with a particular orthographic type; of instantiating certain predicate, function, or constant symbols; of instantiating such symbols in a particular order; of being a sentence; of having two instances of the same syntactic type (e.g. a predicate symbol); and so on.

Semantic properties of a sentence (*given* a modalized interpretation) concern reference or denotation (including extension-assignments), satisfaction, and truth-value allotment relative to various worlds, as well as compositional effects that determine these former semantic properties of wholes on the basis of their parts. These would include properties like being such that one's occurrences of a particular predicate symbol have such-and-such an extension-assignment at a world; of being true at a given world; of being such that one's occurrences of a predicate symbol have their denotations at a world drawn from a particular subset of objects at that world; of being such that one's quantifiers range uniformly range at a world over a single subset of objects existing at that world; of being such that the truth of any constituent predication at a world is determined by whether the denotation of a term at a world is within the extension of a predicate at that world; and so on.

However, *base* semantic properties are only those semantic properties that either belong to minimally interpreted constituents or the most elementary compositional processes by which semantic properties of composites are determined by the semantic properties of their parts (including those compositional properties of 'logical' vocabulary like quantifiers or connectives). For example, consider a language containing a predicate symbol "F" and a constant symbol "a", and a modalized interpretation on which "a" denotes a particular ripe macintosh apple *o* and "F" denotes the set of actual red things. Then the following are base semantic properties of the sentence "Fa" on this interpretation: that of being such that, in it, "F" has as its extension assignment at actuality the set of red things; of being such that in it, "a" has *o* as its extension assignment at actuality; and of being such that in it, the truth of "Fa" at actuality depends on whether the extension of "a" belongs to the extension of "F" at actuality. Another semantic property of "Fa" is that it is true at actuality. But this is not a base semantic property, since it is neither a semantic property of a minimal interpreted constituent, nor a property concerning semantic composition.

Base semantic properties may be particular or they may be general. The property of "Fa" of being such that in it o is the extension assignment of "a" at actuality is a particular base semantic property of "Fa" (on the given interpretation). But another base property of "Fa" is being such that in it the extension of "a" at actuality is drawn from a set of objects at the actual world. This latter property is a more general base semantic property of the sentence, but it is a base semantic property all the same.

By design syntactic properties and base semantic properties of a sentence determine or explain all of its semantic properties (a point that will be relevant to the 'in virtue of' relation invoked in our question (Q′) about true validity).

Now that we have some grip on syntactic and base semantic properties, we can begin to look at circumstances where such properties are shared or not. If we consider a given first-order sentence ϕ given *two* distinct modalized interpretations, these interpreted sentences will share some of the linguistic properties I have mentioned but they may not share others. Minimally, these two sentences will share *all* their syntactic properties (that is what it is for them to be the same first-order sentence). But their base semantic properties could diverge.

Consider again our sentence "Fa". Suppose on interpretation $\mathcal{I}_1$ at actuality, "F" is assigned the set of actual red things and "a" is assigned an existing red apple. On interpretation $\mathcal{I}_2$ at actuality, "F" is assigned the set of actual green things and "a" is assigned an existing green apple. These two sentences-on-interpretations share a non-base semantic property (that of being true at actuality). They don't share any particular base semantic properties relative to actuality aside from that concerning composition for predication. But they do share general base semantic properties even relative to actuality: both of them share the property of being such that, in them, F is assigned an extension consisting of a set whose elements are actual objects; in them, a is assigned an extension at actuality consisting of an actually existing object; and so on.

Again, our focus will be on syntactic and base semantics properties (both particular and general). With that delimitation, and an understanding of when and where such properties are shared, we can finally pick out the particular linguistic properties relevant to (Q′) as those that are *shared* across modalized interpretations.

> The *modalized first order form* (MFOF) of a first-order sentence ϕ consists in the set of syntactic and base semantic properties

shared by ϕ on all of its modalized interpretations.

We can also, by extension, speak of a sentence-on-a-modalized-interpretation bearing its MFOF (though it bears this form trivially, in virtue of receiving a modalized interpretation).

The talk of 'form' here is meant to reflect the ways in which MFOF can be shared by a sentence-type on different interpretations. "Fa"-on-$\mathcal{I}_1$ and "Fa"-on-$\mathcal{I}_2$ share their MFOF. Even though reflect different truths at actuality—that one apple is red and a distinct apple is green—they have some important features in common: they share their broadly model-theoretic mode of semantic interpretation. To break this commonality they would have to be assigned semantic properties in a substantially different manner.

I hesitate to call this notion one of 'logical form' since it would not answer to many preconceptions of that concept. In connection with this, the notion of form here should be distinguished from several notions that have gone under the heading of "logical form". For example, it shares very little with a notion of logical form employed in natural language interpretation to label a representation of structure, sometimes departing from surface grammar, at which all features relevant to semantic interpretation are captured.[6] Also the notion should be distinguished from a notion of logical form on which two sentences share a relevant form if they are both instances of an abstract logical schema.[7] On this usage "$Fa \vee \neg Fa$" and "$Ga \vee \neg Ga$" could share a form for both being instances of the logical schema $\phi \vee \neg \phi$. These sentences would not share MFOF on my usage, since they do not share their syntactic properties.[8]

The importance of MFOF is not that it conforms to some preexisting conception of logical form, but that it exhibits characteristic logical *processes of*

[6] See, e.g., May (1985).

[7] See, e.g., the use of "logical form" in Quine (1970/86, 12,51–2).

[8] The differences between MFOF and form as captured by instantiation from schemas shouldn't be overstated. It would be possible to try to recover something similar to the schematic conception on the strategy I am using by being a bit more selective about the use of syntactic properties (or perhaps syntactic-cum-semantic properties) in a characterization of something like modalized form. This could well be an improvement over my characterization for getting a more minimal and revealing basis for the grounds of the expression of necessities by sentences given modalized interpretations. I don't pursue this further here for two reasons. First, my main goal in this chapter is to show that model-theoretic validity tracks *at least one* 'true' form of validity on the inferential conception. For these purposes MFOF-validity will suffice. Second, as will be very clear soon, I want to highlight the indispensable work that *semantic* commonalities are doing any of the forms of validity—the work that I care about in investigating inference. MFOF highlights these commonalities.

abstraction—processes alluded to in my informal characterization of logical methods from Chapter 2. It reveals one way of privileging certain linguistic properties over others, and thereby (as we will see) privileging one subclass of good inferences over the total range of good inferences in a way that could be revealingly conducive to formalization. In this sense, they provide us with one 'true' sense of validity.

> A first-order sentence type ϕ is *MFOF-valid* iff necessarily, on any interpretation of ϕ that gives ϕ modalized first-order form, ϕ expresses a necessary truth.[9,10]

We can informally argue that model-theoretic validity coincides with this particular notion of true validity as follows.[11]

> Modal-Theoretic Validity Tracks MFOF-Validity.
>
> Let ϕ be a sentence of a first-order language $\mathcal{L}$. Then:
>
> > ϕ is MFOF-valid if and only if $\models \phi$.
>
> For the left-to-right direction, pick an arbitrary model $\mathcal{M}$ of $\mathcal{L}$. Such a model can be extended to a modalized interpretation $\mathcal{I}$ of $\mathcal{L}$ simply by incorporating arbitrary model-theoretic interpretations of $\mathcal{L}$ in non-actual worlds. (I assume, substantively, that materials for constructing

[9] I have here treated the 'in virtue of' relation relevant to validity as supplied by metaphysical necessitation for simplicity. This is certainly an oversimplification. We would want the properties relevant logical validity to secure the necessary truth of sentences in a manner that *explains* them. Necessitation is a necessary, but insufficient condition on being appropriately explanatory (e.g. because it is a symmetric relation). That said, I've set up the characterization of MFOF (in particular, by focusing on *base* semantic properties) so that it would support asymmetric forms of explanation. I don't have the space here to adequately explore the space of options for the particular explanatory relation which would be of greatest interest (and, to be honest, my suspicion is several different explanatory relations might do). It suffices for now that the properties picked out intuitively do the explaining. So I rest content with an intuitive appeal to the 'in virtue of' relation for informal purposes, and necessitation in somewhat more formal contexts like this one.

[10] Here I attribute validity to first-order sentence types, whether or not they are interpreted. The thought is that sentences can be attributed the *conditional* property of MFOF-validity regardless of whether they are interpreted. There are reasons to be interested in only interpreted objects of validity that actually *have* the properties conditionally attributed—an issue which I'll return to consider in Chapter 9.

[11] The basic idea of the argument to follow can be found in McGee (1992). See also Shapiro (1998), Sagi (2014).

such models always exist, since abstracta such as numbers and sets exist necessarily.) By construction, $\mathcal{I}$ gives ϕ modalized first-order form and gives the same truth-value assignments to sentences of $\mathcal{L}$ at the actual world as $\mathcal{M}$. By hypothesis ϕ, on $\mathcal{I}$, expresses a necessary truth, and so is true at the actual world on $\mathcal{I}$. It follows that ϕ is true in $\mathcal{M}$.

For the right-to-left direction, we show the contrapositive. Suppose there is a metaphysically possible world w such that there is an interpretation $\mathcal{I}$ at w giving ϕ modalized first-order form, but on which ϕ does not express a necessity on $\mathcal{I}$. So there is some (perhaps distinct) w' at which ϕ evaluates to falsity on $\mathcal{I}$. In bearing modalized first-order form at w on $\mathcal{I}$, ϕ at w' on $\mathcal{I}$ receives base semantic properties determinative of its truth at w' by model-theoretic means. So there is some model $\mathcal{M}_{w'}$ with objects drawn from w' such that $\mathcal{M}_{w'} \not\models \phi$. Since the domain of $\mathcal{M}_{w'}$ is set-sized, there is a bijection f (existing at w') between the domain of $\mathcal{M}_{w'}$ and some arbitrary equinumerous set of sets. Using that bijection, we can define a model $\mathcal{M}'$ that is isomorphic to $\mathcal{M}_{w'}$ in the natural way. (For example, for any objects $o_1, \ldots o_n$, let $f(o_1), \ldots (o_n)$ be in the extension of a predicate symbol F in $\mathcal{M}'$ just in case $o_1, \ldots o_n$ is in the extension of a predicate symbol F in $\mathcal{M}_{w'}$. And so on.) ϕ's falsity in $\mathcal{M}'$ is preserved by isomorphism. Since the domain of $\mathcal{M}'$ consists only of sets, $\mathcal{M}'$ is a model of $\mathcal{L}$ at the actual world that falsifies ϕ. So ϕ is not true on all models.

The intuitive idea behind the left-to-right direction: When ϕ expresses a necessity in virtue of its modalized first-order form, its truth at actuality is secured by an assignment of properties that is given by first-order model-theoretic means—it's just that such a model only contributes the properties relevant to ϕ's truth at that one world. And because each model can be extended to a modalized interpretation, that is enough to show that a guarantee of truth at actuality given *any* modalized first-order form will translate to truth in any model.

The intuitive idea behind the right-to-left direction: First-order logic is expressively weak. As a result, when a world falsifies what is expressed by a first-order sentence, not much of the world's structure is needed to do this. And by contrast, given the plurality of sets, actual interpretations can be richly structured. In particular, we can always find set-theoretic structures for such

actual interpretations that mirror the structure of any possibility relevant to the falsification of what is expressed by a first-order sentence.

The foregoing argument can be extended to show the equivalence of first-order model-theoretic consequence and MFOF-consequence with obvious changes. Together these show that there are *some* ranges of discourse over which model-theoretic validity and consequence at least extensionally tracks a 'true' form of validity and consequence. This is *some* validation of the importance of model-theoretic consequence. It shows there are certain possible kinds of discourses where model-theoretic consequence directly tracks a 'true' form of logical consequence on the inferential conception.

That said, this argument leaves open an important question. First-order validities are those that express necessities in virtue of some specific batch of linguistic properties. But *just how prevalent are those linguistic properties*? That is, to what extent do we find discourses with sentences bearing modalized first-order form? (More cautiously put if we are interested in natural language correlates of good inference: to what extent do we find natural language discourse that is regimentable without distortion of truth-conditions into first-order sentences bearing modalized first-order form?)

This turns out to be a significant and complex question worth probing in some detail. To preview: modalized first-order form comprises a highly specialized set of semantic properties. The sum of these properties arises clearly and naturally within certain mathematical settings and discourse about certain abstract objects. But discourse about other subject-matter, even once superficially regimented in a first-order language, seems very unlikely to exhibit all relevant properties while remaining relatively faithful to the semantics of the original discourse. How this affects the applicability of first-order model-theoretic consequence will depend on how relaxing certain features of modalized first-order form affects the truth-conditional contents expressed by first-order sentences. And *that* often simply remains a highly contested matter.

To explain these ideas it will be helpful to factor modalized first-order form into three components: the properties it imposes on (constant, function, predicate, and variable) denotations, the properties it imposes on quantification, and the properties it imposes on non-quantificational compositional processes (including those associated with connectives). This will allow us to look in turn at whether, and how, discourses could fail to instantiate each subset of properties.

> The *(modalized first-order) denotational form* of a first-order sentence ϕ consists in the base semantic properties imposed on constant, function, predicate, and variable denotations in ϕ on all of its modalized interpretations (e.g., that every constant symbol is assigned a referent at a world drawn from the domain at a world, that every predicate symbol is assigned an extension drawn from a subset of the domain at a world, etc.) are also in force for ϕ at w on $\mathcal{I}$.
>
> The *(modalized first-order) quantificational form* of a first-order sentence ϕ consists in the base semantic properties imposed on quantifiers in ϕ on all of its modalized interpretations (those properties relevant to their behavior as first-order unary restricted quantifiers ranging uniformly over a non-empty set of objects that exist at each world).
>
> The *(modalized first-order) compositional form* of a first-order sentence ϕ consists in the base semantic properties on compositional processes for non-quantifiers in ϕ on all of its modalized interpretations (that predication satisfaction is determined by predicate and term denotation at a world in the standard manner, and that connectives inherit truth-conditions classically at each world, etc.)

A interpreted first-order sentence ϕ bears modalized first-order form if it has modalized first-order denotational, quantificational, and compositional form. This is because a sentence receives its truth-value at every world by essentially standard model-theoretic processes just in case that value is determined at every world in the standard model-theoretic way (for both quantifiers and non-quantifiers) from model-theoretically acceptable denotations.

I will have little here to say about compositional form. Provided other properties constituting modalized first-order form are satisfied by a sentence, I see no special grounds to worry that the compositional processes of model-theory substantially distort the compositional processes whereby truth-conditional content is determined in inferences expressed by sentences that are typical candidates for first-order regimentation (e.g., those involving no modal operators, etc.).

This is not to say that the compositional processes involved in model-theoretic interpretation are precisely those involved in the relevant discourses that are candidates for regimentation. For example, natural language use of the expressions regimented into first-order logical constants (e.g., English "and",

"or") may operate under a general compositional mechanism of functional application,[12] whereas the compositional conditions on first-order logical constants are typically stipulated in separate clauses without any recourse to functional application. Rather, what I am claiming is that the truth-tables for $\neg$, $\wedge$, and $\vee$ plausibly mirror how compositional processes for their natural language equivalents end up determining truth-conditions for composite expressions from their parts—whether they do this by functional application or some other means. Or, at least, this seems plausible for certain important ranges of discourse (for example, those that haven't been proposed for appropriation by dynamic semantic theories). The material conditional $\supset$ is a special case, since it is highly controversial whether natural language indicative conditionals have close semantic connections to the material conditional.[13] But this merely means that we need to be careful to exclude from regimentation in first-order logic any discourse whose compositional operations can't be captured in simple truth-functional terms. And there are many forms of discourse which plausibly remain once such exclusions are made. Additionally these compositional processes might be more controversial outside the context of *bivalence*. But that is more an issue for denotational form than for compositional form.

The properties subsumed under quantificational form, by contrast, begin to raise more interesting issues and to substantially restrict the discourses that could bear something like modalized first-order form. Let me comment on five features of modalized first-order quantificational form in turn, focusing especially on the impact they might have on restricting our ability to regiment natural language sentences into first-order logic while preserving their truth-conditional content.

(a) *Quantifiers are unary operators.*

Natural language familiarly makes use of restricted binary generalized quantifiers (Barwise & Cooper (1981), Keenan & Westerstahl (2011)). But, also familiarly, universal and existential binary quantification are perfectly mimicked by the familiar unary first-order quantifiers $\forall$, $\exists$ with the help of logical connectives in the two-valued setting. The equivalence does fall away in the ternary setting, but this is again more fundamentally an issue for denotational form, which we will come to soon.

[12] See, e.g., Heim & Kratzer (1998).

[13] For an overview of relevant issues, see Edgington (2020).

(b) *Quantifiers have set-sized (or otherwise restricted) domains at every world.*

First-order quantifiers have arbitrarily small uniform domain restrictions (though non-empty ones—on which more further below). The restricted character of first-order quantifiers is a feature very often shared by ordinary language quantifiers, owing to the fact that natural language quantifier domains tend to be restricted by context.[14] The upper bound placed on quantifier domain size by defining logical validity as truth in all models with *set-sized* domains raises a separate worry that we may fail to capture the logic of unrestricted uses of quantifiers, whose domain cannot be the members of a set. It is already controversial whether there ever is truly unrestricted quantification.[15] But even if there are unrestricted quantifiers, the 'squeezing argument' of Kreisel (1967) can be used to (informally) argue that MFOF-validity, relaxed to allow for the possibility of quantification over classes, or over 'everything', will continue to be tracked by first-order model-theoretic validity (see Appendix B for details).

(c) *Intra-world, domains are uniformly applied to all quantifiers.*

Quantifier domains of ordinary language are highly sensitive to context,[16] even intrasententially. But there are certainly limited ranges of discourse where either contextual contributions to domain restriction are constant, or shifts in context are irrelevant to quantifiers' contributions to truth-conditional content. Again, this merely requires us to take some care in selecting discourses relative to which first-order principles will have application.

(d) *There are no substantive inter-world requirements on quantifier domains.*

When I defined modalized first-order form, for simplicity I did not stipulate any connections between quantifier domains at different worlds. But we can, if we like, reintroduce many reasonable constraints of that kind—altering the notion of 'form' at issue—while preserving the argu-

[14] von Fintel (1994), Stanley & Szabó (2000).

[15] E.g., see Dummett (1991, 1993), Boolos (1993), Cartwright (1994), Williamson (2003), and the articles in Rayo & Uzquiano (2006).

[16] Again, see von Fintel (1994), Stanley & Szabó (2000).

ment structure.[17]

The left-to-right direction of the argument only requires that whatever the relevant form is, a first-order model at a given world can be arbitrarily extended to a modalized interpretation which exhibits the relevant form. That should be a minimal constraint on *any* way of generalizing the model-theoretic apparatus to give first-order sentences truth-conditions.

The right-to-left direction requires a related feature: the ability to take some kind of interpretation of a sentence ϕ (giving it truth-conditions) on which it satisfies the relevant conditions of 'form' at a world w (whatever those may be) and use it to construct a model which doesn't alter ϕ's semantic features at w. This is where introducing connections between quantifier domains at various worlds in any redefinition of modalized first-order form could raise trouble. Such connections could prevent ϕ from having its properties assigned at some worlds by standard first-order model-theoretic means.

For example, suppose quantifiers of a first-order sentence on a the new form of interpretation range, for every w, over the intersection of some set S (e.g. some set of actual existents) and the objects existing at w. This might allow the domain of quantification to be empty at some worlds but not others. Of course, no model has an empty domain, so this would mean that the truth-conditional properties of the sentence relative to some worlds may not be assigned by first-order model-theoretic processes.

So we can introduce constraints on connections between quantifiers in our characterization of modalized first-order form, except those that would prevent sentences with that form from having their properties assigned at all worlds by model-theoretic means. As concerns quantifier domains, this merely means that quantifier domains must always be non-empty and set-sized. But, as noted above, the requirement of being no larger than a set in size doesn't actually present any serious obstacle for the argument. So the only real constraint is that quantifier domains must always be non-empty. As long as we preserve that feature, the left-

[17] Though we may need to alter the characterization of a 'model' accordingly.

to-right direction of the argument can continue to go through.[18] That said, this constraint is worth of independent consideration.

(e) *Quantifier domains are non-empty at every world.*

Natural language quantifiers appear to carry a linguistic presupposition that their domains are non-empty.[19] But the assumption made by model-theoretic interpretation is *much* stronger than this: it is the mere interpretation of quantification, not the truth (or felicity) of any quantified statement, that requires the existence of an object. Modalized first-order form, in generalizing model-theoretic interpretation, requires an existent at each metaphysically possible world. With this requirement, we are making new *kind* of commitment that merits commentary.

Thought of broadly, semantic properties of expressions are, or determine, relations between those expressions and the world. Insofar as they concern reality semantic properties can, when instantiated, impose

[18] Gómez-Torrente (2008) has argued that if first-order quantifiers range rigidly over some set S of actual existents—i.e. quantify over S even at worlds where some, or even no, elements of S exist—and we accommodate a logical predicate E which, at a world, is true of all and only the things existing at that world, then we will have examples of model-theoretically valid sentences which are not necessary. A simple example might be $(\exists x)(Ex)$. This is true in all models, since all model domains are non-empty. But it may express a content false at some worlds, given Gómez-Torrente's assumptions, since the objects of the quantifier's domain at the actual world may not exist at some counterfactual world. Relative to that counterfactual world w, $(\exists x)(Ex)$ is true just in case some object from the *actual* quantifier domain exists at w—and that may not hold.

If we treat rigidity of quantifier domains and an existence predicate as part of a sentence's form, then this example also shows that logical validity in my sense, *relative to that particular choice of logical properties*, will cease to coincide with model-theoretic validity. It is worth noting that the failed equivalence here does not merely owe to the interpretation of quantification, though. For the reasons just given, rigid quantification alone won't disrupt the argument for equivalence.

I accept the failure of equivalence between model-theoretic validity, and necessity in virtue of the relevant form, granting the treatment of an existence predicate as logical. Indeed, I accept the possibility of much simpler failures of equivalence from predicates that are traditionally not treated as 'logical,' since I am open to the view that logicality is largely stipulative. An example of such a failure (when treating the predicate "red" as logical) is given below—though an epistemic logic with "knows" treated as logical might provide a less contrived example. The grounds for the failure of equivalence for, say, empirical predicates treated as logical will be identical to the grounds for the failure of equivalence when an existence predicates is treated as logical: model-theoretic methods tend to fail in the context of logical vocabulary with non-constant intensions. Again, see further below for more discussion of this point.

[19] See Strawson (1952), Karttunen (1973), Gazdar (1979), Soames (1982), van der Sandt (1988), and Zeevat (1992).

more or less stringent requirements on the world. For example, as I'll discuss shortly, perhaps the semantic property of referring to an object requires an object to exist. Or perhaps it requires the weaker fact that there be a possible existent.

Requiring quantifiers to range over a non-empty domain either just at the actual world, or at all possible worlds, is to give them a semantic property which I suspect many will feel places a 'substantive constraint' on how the world is. And we can see this idea seep into the characterization of validity.

For example, $(\exists x)(x = x)$ is model-theoretically valid. It is also an MFOF-validity. But it is important to know why the latter holds. To say it is an MFOF-validity is to say the following: at any world in which which $(\exists x)(x = x)$ is given the linguistic properties of modalized first-order form, including the property that the domain of its quantifier ranges over a non-empty domain at every world, $(\exists x)(x = x)$ expresses a necessary truth. Note that $(\exists x)(x = x)$, qua uninterpreted first-order sentence-type, would have been valid in this sense *even if* (perhaps *per impossibile*) it were metaphysically possible for nothing to exist. Even in this case, $(\exists x)(x = x)$ would continue to be MFOF-valid trivially, since it could never have modalized first-order form. In fact, *every first-order sentence with a quantifier would be trivially MFOF-valid*. If it were possible for nothing to exist, having modalized first-order form would require the attribution of a semantic property which no quantified expression could possibly have.

Also $(\exists x)(x = x)$ would not be MFOF-valid if we allowed restricted domains to be empty at a world. So the fact that $(\exists x)(x = x)$ is MFOF-valid reflects a choice: a choice to focus on semantic properties which an expression could only ever have if the existence of at least one object is a metaphysically necessary truth. I personally think this *is* a necessary truth. But it is worth noting that the MFOF-validity of $(\exists x)(x = x)$ is not of itself capturing that fact. Validity in virtue of a set of linguistic properties, as I've defined it, can be held trivially if the properties are uninstantiable. The MFOF-validity of $(\exists x)(x = x)$ is only ever non-trivially witnessed, with expressions possessing the relevant modalized first-order form, on the prior assumption of the necessity of some ex-

> istent or other. Logic could never *tell* us something must exist. It can only enshrine it as a framework presupposition.
>
> I think that as long as we bear this caveat in mind, there is nothing problematic about using quantification in this way. But it may further restrict the kinds of discourse which are fruitfully regimented into first-order languages for the application of first-order validity. This is especially true given that first-order techniques are modeling systematically restricted domains. In ordinary language, such domains are restricted by something like a property of objects. What we should require of a discourse uniformly restricted by such a property to be legitimately regimented into first-order logic is that *the property is necessarily instantiated*. Arguably, only a relatively restricted range of properties would fall under that heading.

On the whole, the constraints imposed by modalized first-order form on quantification are significant, but leave open broad ranges of discourse that can be modeled in first-order terms. The constraints require modeled quantified discourse to be (a) uniformly restricted (b) over objects satisfying a necessarily instantiated property. Discourse uniformly about abstract entities, to take a salient example, can often easily satisfy conditions (a) and (b).

So let's turn to the key aspects of denotational form, which are that constant expressions necessarily refer, and applications of predicate symbols are necessarily bivalent.

For the moment, let me set aside the necessitism—the view that everything that exists does so necessarily.[20] If we allow that some objects exist only contingently, there are worries that only select branches of ordinary discourse will have necessarily referring terms. The clearest cases will probably involve abstract objects, since many of these exist necessarily if they exist at all.[21] What

[20] See Williamson (2013) for a prominent defense of necessitism. Familiarly, Williamson's argument depends in part on taking classical logic to apply to ordinary existents. Part of what is being argued here is that, on my conception of logic, this should be extremely controversial as a starting point. This is because on the conception of logic I have on offer, it must take its shape from *independently established* theses in metaphysics or philosophy of language. It cannot be used to *motivate* them (e.g., on the grounds of its simplicity or utility). Of course Williamson openly endorses a very different conception of logic than we are exploring here—one on which logical truths are generalizations of certain sorts. I am not sure this conception does any better at motivating the classical picture for Williamson's purposes, though I cannot explore this question here.

[21] Abstractness does not obviously guarantee necessary existence, though, especially if the

about names that refer to ordinary things like "Obama", the 44th president of the United States who, intuitively, could have never existed? Such cases raise a tricky semantic question: is it possible for a referring expression n that actually refers to an object o to also refer to o at worlds where o doesn't exist? I won't try to settle this question here. But I will note that either way we end up answering that question, we may have to forgo the application of first-order tools. If an expression cannot refer to an object at worlds where it doesn't exist, then terms which refer rigidly in the sense of Kripke (1980) (i.e. refer to that object at any worlds at which it exists, and no other objects at worlds where it doesn't) will not refer at some worlds, violating the relevant constraint on modalized first-order form. If, by contrast, expressions can refer to an object at worlds where it doesn't exist, those selfsame rigid expressions will still violate a constraint on modalized first-order form: they will have a denotation at a world which is not drawn from the domain of things existing at that world. Either way, we could not (without further argument) assume that classical logic applied to discourse involving rigid designation of contingently existing entities.

These are troubles that arise particularly for individual constants, insofar as they model the use of ordinary language names (and insofar as names are rigid, as Kripke maintained). Attempts to model the behavior of definite descriptions will also fail to satisfy the constraint of necessary reference whenever their associated descriptive property is not necessarily satisfied. This again appears to be the typical case outside of abstract settings.

What about the second assumption of denotational modalized form: that applications of predicate symbols are necessarily bivalent? Here we encounter a familiar tangle of theoretical issues. There is a wide range of challenges to bivalent predication. Some of these are local to certain special classes of predicates. Many have argued that treatment of semantic paradoxes calls for truth-value gaps or gluts in the application of semantic terms like "true".[22] But there are also motivations for thinking bivalence fails pervasively. This would occur if truth-value gaps or gluts resulted from vagueness, since virtually all ordinary language predicates are vague.[23] Also, some have argued that semantic anoma-

abstract object is characterized in terms of non-abstract things. For example, it is controversial whether the singleton set containing only Socrates exists necessarily.

[22] For prominent instances of gappy treatment: van Frassen (1968, 1970), Kripke (1975), Field (2008). And for glutty treatment: Priest (1984, 2006).

[23] Supervaluationist treatments of vagueness lead to gaps, though it is worth adding the caveat that classic supervaluationist treatments like that of Fine (1975) and Lewis (1982) moti-

lies, or 'category mistakes', like "the number six is red" leads to failures of bivalence.[24] Though anomaly may arise rarely in ordinary discourse, it is arguably rife in a language considered as comprising the totality of its syntactically acceptable constructions.

All of these cases are controversial. The important thing to note for now is merely that the assumption of bivalent application is non-trivial, and might require either settling or prejudging the aforementioned issues. On the current conception of logic, these issues are *pre-logical*.[25]

Let's pause here and take stock. I began by arguing that truth-in-all-first-order-models tracks necessity in virtue of modalized first-order form—the latter being a true form of validity. I then asked: how pervasive is modalized first-order form, especially if we aim to translate ordinary discourse into a first-order language without warping the truth-conditions of the original discourse? The frustrating answer is that it is highly controversial. Modalized first-order form would only uncontroversially be possessed by first-order regimentations of discourse systematically restricted by a necessarily instantiated property (or systematically unrestricted), in which all rigid reference is restricted to necessary existents, all 'descriptive' (or functional) reference is necessarily satisfied, and bivalent predication is systematically guaranteed.

Now, there is one important branch of discourse which often has all these properties: mathematical discourse. Frequently, discourse in mathematical proof is systematically restricted to a particular set of abstracta (numbers, groups, fields, knots, etc.), referring terms either pick out abstracta (like numbers) that exist necessarily, or descriptively refer with properties necessarily satisfied (by the same abstraction at each world), and all predication, in part because of the domain restriction, escapes worries from paradox, vagueness, and semantic anomaly. These semantic properties may be shared by some other

vate the supervaluationist framework on the basis of massive ambiguity in vague terms. It is a complex matter how the latter proposal interacts with applications of first-order logic, especially if it is meant to model properties of good inference tracked by linguistic forms. For subvaluationist treatments of vagueness leading to gluts see Hyde & Colyvan (2008), Weber (2010), Cobreros (2011).

[24] See Thomason (1972), Lappin (1981), Shaw (2015).

[25] By this I mean they are precursors to the formalized study of inference in non-idealized settings. Obviously these issues are not 'pre-logical' in the sense that we can forgo inference (including inferences sometimes categorized as 'logical') when investigating them. But the justification for and against, say, failures of bivalence due to semantic anomaly *needn't* turn on the acceptability of contested inference rules. See the discussion of Excluded Middle in §7.2 below.

forms of restricted discourse about abstract entities that exist necessarily.

To sum up, we can say the following about classical logic: first-order logic's relevance to actual good deduction is safeguarded at least to some extent through its applicability in mathematical domains. Beyond that, however, its application is subject to many controversial applications of highly stringent semantic constraints. And it is reasonable to worry that the sum of such constraints is rarely satisfied outside anything resembling discourse about mathematics or other abstracta.

If modalized first-order form were indeed rare, would this make applications of first-order logic sparse? This would not immediately follow, even on my conception of logic. It will depend on whether or not constraints on modalized first-order form can be relaxed, while still allowing for an equivalence argument of the form that I gave above. As I noted when discussing interworld constraints on quantifier domains, there is some room to redefine the properties of modalized first-order form while preserving an equivalence argument with only minor alterations. But, as also noted above, sometimes it is clear that there is no extension of the argument to be given, usually because we can find simple and direct counterexamples to the equivalence. For example, this will clearly occur for failures of bivalence if they have anything like the compositional semantic consequences they are typically taken to have.

There may be intermediate cases where it is contested how exactly a semantic feature should influence truth-conditional content. For example, what happens to a predication of a name when the name fails to refer? If the result is a truth-value gap that has an 'infectious' character, this can greatly perturb classical inference rules. If the result is falsity, there is less ground for concern. I think in cases like this, a mix of foundational inquiry in the philosophy of language, and empirical matters in linguistics, should be used to settle the issue.

One might wonder: isn't this just a stipulative matter? Can't we develop the 'simpler' version of logic, in which reference failure leads to as few perturbations as possible? The answer is that tough cases can sometimes can be viewed as room for arbitrary stipulation of linguistic rules. But the problem is that we are typically interested in *our actual inferences*—that is, whether certain inferences we actually make are good, or not. And if that is our concern, it will matter what contents we *actually* think. These are arguably the contents expressed in ordinary language. We can stipulate substantially new ways of using that language—changing the truth-conditional contents expressed by the

sentences of a language—only at the cost of ceasing to track the thoughts we actually think. As such, stipulation preserves simplicity at the risk of ceasing to model the target phenomenon. This is thus one place where inquiry into the foundations of logic can run up against empirical considerations. These empirical considerations matter, because language is our first and best resource for understanding the *actual* contents of our own thoughts. And sometimes, some structural aspects of our thought only become clear to us after that non-trivial empirical linguistic work.

Let me sum up what we can conclude so far.

Is first-order model-theoretic consequence a form of logic, in my sense of logic? The answer is "yes", provided we are interested in the patterns of good inference that are revealed in linguistic fragments that meet the series of stringent semantic constraints encapsulated in modalized first-order form. Mathematical discourse gives us a clear case where the those semantic constraints are jointly satisfied. But given how contested the general application of each of those semantic constraints is, we should be wary of taking first-order logic to give us insight into the operation of good inference beyond mathematical domains.

We might put the lesson here as follows. First-order logic is often prized for being simple and well-behaved. But first-order logic is only well-behaved because it is the logic of semantically well-behaved topics. One could try to argue that all discourse is as semantically unproblematic as mathematical discourse. But when we compare the semantic simplicity modalized first-order form imposes on naming, referring by description, predication, and quantification with the extraordinary complications we find in the literature on these semantic devices as they appear in a natural language like English, we have strong reason to doubt that first-order logic provides us with a broadly applicable model of good inference. This is so even when considering only branches of discourse which can be naturally recast with the syntactic forms of first-order languages.

I want to make one final comment about the relation between model-theoretic validity and 'true' logical validity, as I've characterized it. The case of first-order model theory is interesting, because it shows that we are sometimes able to capture information about metaphysical necessity in virtue of certain classes of linguistic properties indirectly, using only the resources of actuality. It is perfectly possible to think of the models we permute in defining 'truth-

in-all-models' as holding fixed the actual interpretations of 'logical' vocabulary, and merely cycling through various different actual interpretations of the non-logical vocabulary—all the while focusing only on extension-level properties belonging to expressions at the actual world. Even if we were doing this, we would be able to glean modal information about the contents expressible by first-order sentences in the process. The equivalence argument above shows how this is made possible by the expressive simplicity of sentences with modalized first-order form and the comparative interpretive complexity afforded by set-theoretic resources at actuality.

This raises important questions: For what other batches of linguistic properties besides those involved in modalized first-order form can we safely extract modal information in this indirect way? When can focusing on mere reinterpretation at actuality, and holding actual interpretation of 'logical' terms fixed, indirectly give us information about metaphysical necessity in virtue of relevant semantic properties of those terms? Note: this question is different from the one most recently asked, which was: "For what sets of linguistic properties can first-order validity track expression of necessary truth in virtue of possessing those properties?" Now we ask: "For what kinds of linguistic properties can we define validity using variations of actual interpretations, perhaps generating a logic *other* than first-order classical logic, and track necessary truth in virtue of those properties?"

Lamentably, using actuality-focused interpretations to capture, or model, counterfactual modal profiles and the semantic features that ground them only works because of the very specific choice of linguistic properties involved in modalized first-order form. These techniques don't naturally generalize to track "necessary truth-preservation in virtue of linguistic properties L" for variable L.[26] This point, extensively argued in Etchemendy (1990, 2008), can be pressed in various ways. A simple one is to consider an expanded set of linguistic properties. Suppose we ask which first-order sentences express necessary truths in virtue of their modalized first-order form *and* the fact that some empirical predicate—say "red"—means what it does. This question is about a new form of 'true' validity—one relative to an expanded set of linguistic properties. Suppose further that there is some finite number n of red

[26] This doesn't mean that *model-theoretic* techniques can't generalize. Rather: model-theoretic techniques *construed and regulated as focused on actual interpretation* cannot do this. See the discussion of different forms of model-theoretic interpretation in §7.2.

things. Then if we employ the model-theoretic definitions of satisfaction and truth while holding the actual interpretation of "red" constant (in the following sense: that as we vary the domain, we always assign an extension to "red" consisting of the actual red things in that domain), then claim that there are no more than n red things will be marked as model-theoretically valid. This is because by hypothesis no subset of the domain of actual existents can contain more than n red things. But of course, it isn't a necessary truth that there are no more than n red things. What this simple case shows is that a technique of holding actual interpretations of 'logical' terms fixed while permuting actual interpretations to get a mix of information about modal profiles and semantic grounds can sometimes work, but only because we didn't ask about truth in virtue of linguistic properties that included the property of sentences that, in them, "red" refers to red things. Similar problems will arise, for example, if we apply those techniques to epistemic logics containing predicates like "knows" or "believes". (And, familiarly, epistemic logics are not investigated in that way.)

Where does my above equivalence argument fail in these cases? Trouble will arise in defining the isomorphism in the right-to-left direction. On that direction of the argument, we took a hypothesized world which falsifies the content expressed by a sentence ϕ with modalized first-order form, and used it to construct a model at the actual world that falsified ϕ. In the case just given, our language contains a predicate "red", which on its modalized interpretation denotes at each world the set of red things at that world. And "red" is now also being treated as 'logical,' so (on the actuality-focused use of the model-theoretic apparatus) that we cannot change the fact that "red" applies only to actual red things in model-theoretic interpretation. Given a hypothesized world w with $n + 1$ red things falsifying the modalized content of ϕ, we are guaranteed to fail to produce a 'model' at the actual world isomorphic at w to the model which mirrors the way ϕ's semantic properties are determined (at w) on its modalized interpretation. The problem arises at the base level of constructing the isomorphism, owing to the predicate "red": no subset of objects at the actual world contains n+1 red things.

A key part of the failure of recapturing modal information from a permutation of actual interpretations in this instance appears to be the fact that the predicate "red" has a non-constant intension (in the sense of not applying to the same set at each world—as opposed to not denoting the same property at

each world).[27] Analogous failures for a knowledge or belief predicate will arise for roughly the same reasons.

This is hardly an objection to the model-theoretic techniques. It is only to note that if we are interested in how various linguistic properties necessitate necessary truth, then the techniques employed in first-order model theory must be extended with care when it is used to track true forms of validity for ranges of properties other than those involved in first-order modalized form. Perhaps, of course, that batch of properties is special. One could argue that they are the 'truly logical' linguistic properties. As I've flagged, I will not discuss the question of whether there are 'true' logical properties here. I will merely note that even if there are linguistic properties that are privileged in *some sense*, there could be no objection to an investigation of the conditions of good deduction that considers formalized systems of inference based on properties that are not privileged in that sense. That is, if we start from the idea that we are investigating patterns of good deduction, there is no clear reason to limit ourselves to some particular subclass of them. Whatever reason we had for limiting ourselves in this way does not seem like it could stem from the theoretical goals with which I began this book.

The recent point about the failed generalizability of model-theoretic techniques construed as permuting actual interpretations raises to salience the topic I would like to turn to next. Namely: to what extent can model-theoretic techniques contribute to something like an analysis, reduction, or explication of logical validity—even if only validity relative to modalized first-order form? Worries about properly generalizing model-theoretic techniques certainly seems to call this idea into question.

7.2 A Reduction of Consequence

In Chapter 2, I gave an explanation of logical truth in terms of deductive inferential goodness. We saw that a necessary condition on inferential goodness is the preservation of truth across a range of possibilities. Logic studies that aspect of good inference indirectly, by investigating how necessary truth-preserving relations arise in the sentences of a language in virtue of some subset of their linguistic properties. Though I initially left open what kinds of worlds

[27] It is not the whole of the story, though. Identity also has a non-constant intension. So would an existence predicate. Neither (on their own) creates trouble for the equivalence argument.

a deductive inference must preserve truth over, in Chapters 4–5, I filled in that gap. Deductive inference involves a cognitive act of crowding-out which resolves inquiry into a question. And acts of crowding-out were in turn found to have constitutive connections to metaphysical possibility. As such, what good deductive inference requires was found to be a metaphysically necessary truth-preserving transition.

In giving the foregoing analyses of logical consequence in terms of properties relevant to good inference, and of inference in terms of crowding-out relations, I claim to have supplied a reduction of logical consequence in throughly non-logical terms. In the simplest form, of course, we have reduced questions about 'true' validity and consequence to questions about necessitation relations among contents that hold in virtue of certain linguistic properties. But, of perhaps greater importance, by embedding that enquiry within a broader project of investigating the mental activity of inference, we have given increased clarity to the shape and purpose of logical investigation that can give guidance when we ask questions about particular logical frameworks as well as broader foundational questions about them.

On the resulting view, questions about a logical formalism will reduce to questions in the philosophy of of mind (concerning the nature of the mental event of deductive inference, and the crowding-out relations which help constitute it), in the philosophy of language and linguistics (concerning the nature of semantic properties, and how they determine the truth-conditional content that can figure as the objects of the acceptance states involved in inference), and metaphysics (concerning the nature and extent of metaphysical modal space in its role of shaping metaphysically necessary truth-preservation). Though we may use inferences—among them 'logical inferences'—in debating what theses hold in these branches of inquiry, the questions in them which most directly shape the character of logical investigation are not themselves logical, and will often be settled independently of contested logical inference rules. Or so I will argue.

Not only this, but first-order model-theoretic techniques in particular can now be understood as a helpful tool in that logical investigation. If we want to explore a consequence relation of metaphysically necessary truth-preservation among the contents of several sentences, which holds in virtue of their modalized first-order form, the natural way to do this is to model two things: changes in the ways the world might be (to investigate whether the contents necessarily

preserve truth), and changes in the interpretation of the properties that aren't part of modalized first-order form (to investigate whether the necessity holds 'in virtue' of the remaining properties).

What is intriguing about first-order models is that they could be viewed as varying these two dimensions using a single formal mechanism. Altering a model-theoretic interpretation involves a disjunctive change in either 'non-logical' linguistic properties or a change to how the world might be (or both). The equivalence argument of §7.1 again can show how shifts in model-theoretic interpretation can safely play both roles at the same time. This is my preferred way of understanding the role of model-theoretic techniques and the one most consonant with the aims of the inferential conception of logic. What may not be obvious is that it is a slightly different *construal* of the way that a model 'models' than that just discussed at the end of §7.1—it is one in which the model is thought of as at least in part 'directly' modeling modal information rather than *merely* varying actual interpretations (even if to capture modal information indirectly). We'll discuss this idea more very shortly.

I want to situate my proposed reduction of logical consequence, and its attendant construal of the importance first-order model-theory, against the backdrop of a core line of argument pressed by Etchemendy (1990, 2008) against the idea that model-theoretic techniques could be, or be part of, a proper analysis of logical consequence into non-logical notions.[28] His idea is that model-theoretic techniques could at best capture 'intuitive' logical consequence as an extensional matter. But they could not provide a useful reduction of logical notions to non-logical ones.

At the heart of Etchemendy's argument is the idea that logical truth and consequence have some epistemically, modally, or semantically privileged status—what I will call, following Sánchez-Miguel (1992), the 'Modal Knot.' For example, logical truths may be *a priori*, or analytic, or necessary (on some modality), or some combination of these. What exactly the privileged status is matters less than that there is some such status, and that it is what gives logic its importance. For example, it is what allows us to expand our knowledge by logical deduction, know logical truths without investigation, and so forth. The point is that any suitable analysis or reduction of logical notions should capture a special status of roughly this kind.

[28] For related concerns about analysis or reduction see Pap (1958), Kneal & Kneale (1962), Field (1989).

As I see it, Etchemendy presses a dilemma. There are two salient ways to construe first-order models (to be elaborated shortly): interpretationally and representationally. On the first, interpretational construal, we fail to have anything resembling conceptual equivalence for failure to explain any aspect of the Modal Knot. By contrast, on the representational construal, we have the potential for a kind of conceptual equivalence, but it is one which simply presupposes logical notions, and cannot explain them. Neither way will we have a reduction of logical validity to non-logical notions.

The two construals of first-order models just alluded to—interpretational, and representational—can be understood as two ways of interpreting the role played by varying those models in ascertaining logical validity. To preview: my conception of logic falls into the second, representationalist category. But to understand the force of Etchemendy's dilemma, it will still be important to review its first, interpretationalist horn.

On the interpretational construal, roughly, models vary the extension-level semantic properties (e.g. referent, truth-value) to non-logical expressions, respecting the following constraint: the extension-level semantic property must be one an expression of the same semantic type (e.g. unary predicate, constant term) could have *at the actual world*.[29] Note 'could' is not a form of metaphysical possibility but something more like epistemic or informational possibility: to say a semantic property is 'possible' for an expression to have means that its possession of the property is compatible merely with the information that the non-logical expression is of a certain semantic type.

Such a construal of model-theoretic techniques seems to be motivated by an 'analysis' of logical truth along the following lines:

> A sentence S is logically true iff S is *actually* true, no matter how the non-logical vocabulary in S is interpreted.

Etchemendy notes that if this were a conceptual analysis, it would be a remarkably powerful one. It reduces whatever concepts figure in the Modal Knot associated with logical truth to simple actual truth (via actual satisfaction) and the specification of a class of interpretational variants of S. But the specification of the latter class requires no contested modal, semantic, or epistemic

[29] The formulation is rough here for several reasons. We need to build in the additional constraints discussed in the previous section, such as that constant expressions refer, predicates apply bivalently, and so on. And special conditions are required on quantification.

notions. Accordingly, logical truth is effectively reduced to membership in a pattern of actual truths.

But the attractiveness of the analysis, Etchemendy claims, is precisely its undoing. A key problem is that we can't retrieve aspects of the Modal Knot in the analysandum responsible for logic's privileged role from the weak resources of the analysans. I won't be able to do justice to Etchemendy's entire case here. But it may help to sketch one of his central arguments, which involves a worry about over-generation connected with the point discussed at the end of the last section. The claim that there are no more than n red things may remain a truth no matter how we reinterpret terms other than first-order logical vocabulary and "red", simply because there are actually no more than n red things. And the claim that there are no more than n red things is, if true, contingent, synthetic, and *a posteriori*. It has no privileged status whatsoever. Now, as it happens, perhaps there aren't a finite number of red things. If so, then there will not be a counterexample of the above form involving "red". But even if there are an infinite number of red things, the fact that there are so many red things shouldn't be presupposed as a logical given. That is, the fact that there are an infinite number of red things should not be a presupposition required for an analysis of logical consequence to get its verdicts right.

A tempting reply to the worry raised by this example is that perhaps the concern only arose because we applied the techniques to 'non-logical' vocabulary like "red". But Etchemendy is concerned that the over-generation worry persists even when we restrict attention to first-order resources. Consider the following inference, in the context of taking negation, conjunction, and the material conditional to be logical.

$$P(a) \wedge Q(a)$$
$$\neg P(b)$$
$$\text{So } P(c) \supset Q(c)$$

This inference is not valid in any intuitive sense. But it is only not valid on the interpretational construal if we can interpret the premises to be true while the conclusion is false. And, Etchemendy notes, we can only do this if there are more than two things, and that plurality of things does not fall into two indistinguishable classes. As it happens, there *are* more than two things, and these things do not fall into only one or two indistinguishable classes. But, Etchemendy worries, just as the fact that there are an infinite number of red

things shouldn't be a logical given, neither should the claim that there are more than two distinguishable objects. Even if this claim is a metaphysically necessary truth, it is not a logical truth, and should not be a precondition for an analysis of logical consequence to get its verdicts right. If so, we get the right results on the interpretationalist construal of model theory only by taking on board non-logical assumptions about actual truths. Also, given this, what assurance do we have that the analysans doesn't over-generate relative to other vocabulary instead? And even if it doesn't over-generate, doesn't the mere worry that it could over-generate reveal that we at best have an extensional, and not conceptual, characterization of logical consequence?[30]

This hardly represents the full extent of Etchemendy's concerns for the interpretationalist tradition, and even this argument is incomplete. But I don't want to consider the ensuing dialectic here. This is because the explication of logical truth I've been offering in this book, and the associated construal of the model-theoretic machinery, is not given along interpretationalist lines. My characterization of logical truth, which is offered in a reductionist spirit, openly adverts to metaphysical modal notions twice-over. Recall that on this account, a sentence S is a logical validity, relative to linguistic properties L, just in case, necessarily, for any language containing S in which S bears the properties in L, S expresses a necessary truth. And the construal of model-theoretic techniques that pairs most naturally with this reduction is the one I sketched earlier in this section, on which varying models captures relevant modal and semantic information disjunctively.

To this extent, my proposed analysis comes closer to what Etchemendy calls a "representationalist" construal of model-theoretic machinery. On this view, the varying interpretations in model-theoretic analysis are "meant to be mathematical models of logically possible ways the world, or relevant portions of the world might be, or might have been."[31] Of course we can't construe model-theoretic interpretations as modeling ways things could have been consistently with expressions retaining their actual meanings. For example, such interpretations sometimes assign non-logical constant expressions (e.g., "o")

[30] As it happens, Etchemendy thinks over-generation won't occur if standard first-order logical vocabulary is singled out, and moreover that this can be proved. The problem is that the proof requires appeal to theses about logical truth which presuppose elements from the Modal Knot. The thought is that this proof can give some evidence of the extensionality of the analysis, but precisely at the cost of revealing its conceptual inadequacy.

[31] Etchemendy (2008, 287).

referents they have at no possibilities (e.g. the number 1, or a dog). So the space of interpretations should represent alternative possible ways things could have been, allowing that the meaning of some vocabulary—the non-logical vocabulary—can shift. The refined idea, as Etchemendy puts it, is that "[e]ach model is meant to represent a semantically relevant circumstance…for *some* interpretations of the expressions [treated non-logically]."[32]

This characterization of Etchemendy's fits right into the picture I gave at the outset of this section. To devise an interpretation on which "red" is assigned three members of the domain while that predicate is treated as a logical constant, would be to devise a mathematical model of possible circumstances where there are exactly three red things. To devise an interpretation on which "red" is assigned three members of the domain while that predicate is *not* treated as a logical constant, is to devise a mathematical model on which, for some way of interpreting "red" consistently with its having a predicative meaning (say, the actual meaning of "is a dog"), there are exactly three things that would satisfy that way of interpreting "red" (so, a possible circumstance with exactly three dogs). The point of varying two dimensions—the possibility modeled and a possible way of interpreting expressions—is roughly that we saw arise in §7.1: it allows us to see when meanings of a certain subset of expressions can do the work of securing a sentence's metaphysical necessity.'[33]

This yields a second characterization of logical truth. Again roughly:

> A sentence S is logically true iff S *would be* true in any possible circumstance, no matter how non-logical vocabulary in S is interpreted.

Etchemendy is happy to endorse the representationalist method of interpreting model-theoretic machinery, and its associated characterization of logical consequence. The problem is no longer that we lack a characterization of logical consequence capturing elements of the Modal Knot. Rather, the concern is that our characterization is no longer a reduction, or proper analysis. According to Etchemendy, we will have simply presupposed logical notions in introducing the notion of a 'possible circumstance' in our analysans:

[32] Etchemendy (2008, 289).

[33] See also Hanson (1997), Shapiro (1998) for similar construals of consequence. Critically, neither Hanson nor Shapiro explicitly takes the necessity borne by logical truths to be metaphysical. Shapiro, in fact, characterizes things in terms of 'logical necessity'—which we will see raises precisely the concerns that Etchemendy has for representationalist attempts at reduction.

> …representational semantics give us no such analysis [of logical notions], since the logical notions are used to assess the class of models devised for the semantics. Each model is meant to depict a logical possibility; no two are logically inconsistent; and the "sum" of the models is logically necessary… This has a consequence that some may find disappointing, though it should hardly be surprising. We cannot look to model-theoretic semantics to answer the most basic foundational issues in logic. For example, if we have serious doubts about whether the principle of excluded middle is a logical truth, the classical semantics for propositional languages will not provide answer. For the same intuitions which suggest that it is a logical truth are used in defining the class of structures—truth-value assignments—that are used by this semantics.[34]

The idea, I think, is that when we treat interpretations as varying semantic properties relevant to actual truth-value determination, it is relatively clear what the space of interpretations should look like. But once we view interpretations as modeling possibilities, things will become murkier, and much more controversial. And intuitions about logicality will be informing any view of what the space of possibilities looks like, especially once we see how directly they impinge on contested logical rules.

I think Etchemendy's concerns here have some weight. But I think they are overstated, at least as applied to a broadly representationalist view of the sort of I've been defending so far. In what follows, I want to get clearer on the extent to which my view presupposes logical notions, and the extent to which it can, and cannot, illuminate foundational disputes.

Etchemendy frames his worries for the representationalist analysis by saying that it presupposes a notion of *logical possibility*. There are at least two things Etchemendy could intend logical possibility to be. On the one hand, there is a *proprietary* construal of logical possibility, on which it is construed as a *sui generis* form of modality, typically ranging more broadly than all other forms, including metaphysical possibility. In this sense, it is sometimes claimed that it is 'logically possible' for 2 not to be the sum of 1 and 1, even though it is precluded semantically, metaphysically, and epistemically. On the other hand,

[34] Etchemendy (2008, 294).

Etchemendy may be adopting a kind of *schematic* conception of logical possibility for dialectical purposes. The use of the term would be a means of managing variation in the alternative versions of representationalism about model theory. On this construal, logical possibility is standing in for whatever modal, semantic, or epistemic notions are integral to the operation of logical truth, that end up (allegedly) being captured by the representationalist analysis.[35]

It should be clear that my proposed characterization of logical consequence doesn't appeal to any *sui generis* conception of logical possibility. It reduces logical truth to a combination of linguistic and metaphysical modal facts. This does, of course, fit instead with the alternative *schematic* conception of logical modality just alluded to. To this extent, there is a concern that the analysis doesn't bottom out in a notion anywhere as simple and uncontroversial as actual truth, as we would have had on the interpretationalist conception of logical consequence. But I think we overstate this problem if we think the characterization of consequence can't contribute importantly to foundational

[35] For the record, I'm unsure either of these construals fits very well with Etchemendy's remarks. But I'm also unsure what other modality he could have in mind.

For example, Etchemendy often speaks in ways that conflict with a *sui generis* conception of possibility. He is quite clear that, on his view, there is no privileged set of logical constants, and that we are free to investigate the 'logic' of expressions like "is red" or "is a vixen" just as we are those of "not" or "and". He also sometimes speaks of logical possibilities as "semantically relevant circumstances" that could make a difference to the truth or falsity of a claim (Etchemendy, 2008, 293). The former claim seems to bring us closer to the study of entailment relations in compositional semantic theorizing. And the latter distinguishing theoretical role of logical possibilities is one that is often assigned to metaphysically possible worlds. Each speaks against taking logical possibility to track the very broad construal of *sui generis* logical possibility one finds discussed elsewhere in the literature. And I'm not sure there is another *sui generis* conception of logical possibility to work from.

But other times, Etchemendy makes claims about logical possibility that seem to be inconsistent with treating it as (say) metaphysical possibility. He claims that one virtue of the representationalist construal of model theory is that it explains why we are allowed to vary the size of an interpretation's domain, even when exploring the logic of a language containing unrestricted quantifiers: "First-order structures, viewed as representations of the world, should of course have different domains: this is simply our way of representing the fact that, although the world is the size it is, this could have been different." (Etchemendy, 2008, 291). The "could" here must express logical possibility for Etchemendy's claim to be relevant. But that claim would be highly controversial if logical possibility were then equated with metaphysical possibility. After all, it is a common view not only that there are infinitely many abstracta, but that this is metaphysically necessarily so. Indeed, in a discussion of finitism Etchemendy himself says that he "suspect[s] he agree[s] with" the claim that finitism is necessarily false, so that there are necessarily infinitely many things. (Etchemendy, 2008, 274, n.5) It is unclear which notion of necessity Etchemendy has in mind in acquiescing to this claim, but it seemingly must diverge from the logical necessity appealed to in his discussion of varying quantifier domains.

issues in logic.

To defend this point, I want to discuss two applications of the framework for consequence that I've supplied in which controversial questions in logic are reduced to different questions in non-logical domains.

Let me begin with Etchemendy's own example of Excluded Middle. What would it take for such a principle to fail? It is obvious that we can construct an artificial language, with sentences expressing ordinary bivalent truth conditions, that contains a unary connective $\neg$ and binary connective $\vee$ such that $p \vee \neg p$ is not a logical truth (on a conception of logicality which holds the meanings of $\neg$ and $\vee$ fixed). The concern would be that this artificial language might distort the customary meanings of $\neg$ and $\vee$. Usually, in asking whether the law of excluded middle could fail, we are wondering whether we could have counterexamples to $p \vee \neg p$ with $\vee$ and $\neg$ receiving their *customary* truth-functional behavior in regards to (purely) true or false sentences. But what this means is that to have a genuine counterexample to excluded middle, minimally, there would need to be a third truth-value besides truth and falsity that could somehow characterize *assertoric content* (note: not necessarily mental content). Why assertoric content? Because, on the construal of logic I've provided, these are the only contents that logic can investigate directly with any precision, in its goal of indirectly characterizing relations between mental contents that contribute to good inference. This means that merely devising an abstract three-valued linguistic symbolism is irrelevant to whether Excluded Middle could fail. That symbolism must represent some genuine third status in modeling assertoric content. How could this happen? The two salient options are for the symbolism to model the failure of a sentence to produce assertoric content, or to model a third status beyond truth or falsity that successfully produced assertoric content may have at some world.

Can a sentence with meaningful constituents fail to express assertoric content? This is sometimes argued for on the basis of reference failure.[36] And could a sentence which succeeds in expressing assertoric content fail to be conventionally truth-valued at some world? Several philosophers, such as Dummett (1959) and Glanzberg (2003), have argued that we can't make sense of such a third value. Others, such as Soames (1989, 1999) and myself in Shaw (2014), have argued that on certain conceptions of assertion and assertoric con-

[36] Though for a stronger case based on contingent reference failure, arising from complex demonstratives with unsatisfied nominals, see Glanzberg & Siegel (2006).

tent, we can.

My goal is not to settle this complex issue here. Rather, the critical point for now is that the existing arguments for and against the idea that meaningful sentences can fail to express assertoric content, and those for or against the view that assertoric content can bear a third status, are *not directly based in intuitions about logic*. In particular, neither the arguments of those skeptical of the possibility of a third-status, nor the arguments of those who champion its possibility, antecedently depend on the logical question of whether Excluded Middle is valid. The arguments are based on substantive theories about the nature of assertoric content, how it relates to the semantic values of sentence constituents, what theoretical role assertoric content fulfills, and whether there is any sense in thinking of an abstract tri-partition of possibilities as fulfilling that role. The persuasiveness of either set of arguments does not turn on whether Excluded Middle holds.

Even once these questions are resolved, there are further questions about how failure to express assertoric content, or success in expressing trivalent assertoric content, could bear on mentality, and in particular on inference as a mental act. These questions arise at the intersection of the philosophy of language and the philosophy of mind. For example, there seems to be something confusing about thinking that a contentless sentence could somehow relate to the beginning or end of a mental inference, which inherently operates on content*ful* mental states. Also, even if we make sense of a third value that can be assigned to assertoric content, it will be an important, open question how this bears on the use of such content to characterize mental states.[37] These questions, too, are not questions in logic, but foundational questions in the philosophy of mind and language about the character of mental intentionality, and its relationship to linguistic intentionality. And, again, there is no special reason to think that the logical controversies they will affect, like the question of whether Excluded Middle holds, will be prejudged in deciding them.

Again, my proposed characterization of logical consequence does not settle any of these matters, and so does not settle the question of whether Excluded Middle can fail. But the analysis on offer helps us see when questions in semantic theorizing and in the philosophy of mind impact logic in non-question-begging terms. In this instance, the analysis transforms questions

[37] See Shaw (2014) for a view on which trivalence has different roles to play in the characterization of linguistic and metal content.

about logical principles into substantive theoretical questions not, or at least not typically, settled by a stance on those very principles. And it gives us the tools to see more clearly how the non-logical views in various disciplines impact the shape of logical theorizing.[38]

The case of Excluded Middle is one where my account of logic and inference passes a logical question on to related areas of philosophical inquiry, like the philosophies of language and mind. But debates over logical principles can also find resolution within the details of the account of inference itself. This occurs for Ex Falso—the principle that any claim follows from a contradiction.

Ex Falso is sometimes claimed to be a logical excrescence, accepted dogmatically to maintain the simplicity of a classical logic, when it is plain that it licenses a terrible form of reasoning. Consider Priest's remarks (quoted in MacFarlane (ms/2004)):

> For the notion of validity that comes out of the orthodox account is a strangely perverse one according to which any rule whose conclusion is a logical truth is valid and, conversely, any rule whose premises contain a contradiction is valid. By a process that does not fall far short of indoctrination most logicians have now had their sensibilities dulled to these glaring anomalies. However, this is possible only because logicians have also forgotten that logic is a normative subject: it is supposed to provide an account of correct reasoning. When seen in this light the full force of these absurdities can be appreciated. Anyone who actually reasoned from an arbitrary premise to, e.g., the infinity of prime numbers, would not last long in an undergraduate mathematics course.
>
> (Priest, 1979b, 297)

Since relevantist logicians sometimes press their objection to Ex Falso on the basis of intuitions about good reasoning, MacFarlane (ms/2004) sensibly suggests that to make progress in the debate, we need to get clearer on the normative role of logic. Steinberger (2016) uses MacFarlane's discussion to argue that no plausible bridge principle supports the attempt to justify rejection of Ex Falso on the basis of its being bad reasoning.

I agree with MacFarlane that getting clearer on the normative role of logic is a first step in assessing the relevantist challenge. The problem is that I also

[38] I'll return to consider in much greater detail how this impact might be felt in Chapter 9.

think that getting clearer on the role of logic in good reasoning reveals that questions of normativity are largely irrelevant to the question of whether Ex Falso is valid.

On the view of logic I've been promoting, we have a division of labor between formal logical theorizing and a broader theory of good inference to which it may contribute. And that latter theory of good inference is itself part of an even broader theory of good reasoning. So we in fact have *three* types of questions to ask about Ex Falso:

(1) Is Ex Falso valid (relative to some choice of linguistic properties)?

(2) Do, or can, inferences licensed by Ex Falso count as good inferences?

(3) Do, or can, acts of reasoning by Ex Falso count as instances of good reasoning?

Note that the answers to these questions can in principle come apart. A valid argument form may not model a good inference, since validity is a necessary but insufficient condition on inferential goodness. That was a key result of casting logical investigation in inferential terms in Chapter 2. And a good inference may not be a good form of reasoning. That was a key lesson of Chapter 3: to call an inference good is not to say it ought to be performed, and hence not to pronounce it a good way of adjusting one's beliefs (or other acceptance states). If we frame questions about Ex Falso solely in terms of good reasoning, as relevantists have sometimes done, we are in fact at two removes in the line of questioning from what logic itself should say.

To answer the first question: Ex Falso is valid, relative to any choice of linguistic properties ensuring its key premise expresses a necessary falsehood. That is sufficient to ensure the transition between contents in Ex Falso is necessarily truth-preserving, albeit vacuously. We can emphasize again, that when Ex Falso counts as valid for these reasons, it will count as such because of a theoretically-motivated stipulative restriction of the scope of logic to the study of a necessary but insufficient condition for inferential goodness: metaphysically necessary truth-preservation. Provided inference is really the kind of mental act that I've claimed, we are free to study this relation as the only formally tractable condition contributing to good inference.

Again, this is not to say anything about whether reasoning involving Ex Falso is good reasoning. It is obvious that some, and perhaps all, reasoning

with Ex Falso is bad reasoning. For example, to conclude that one is a deity on the basis of a contradiction one has previously overseen in one's own belief system is mad. But the explanation for why that is a bad form of reasoning is not that it involves an invalid inference, but rather that it involves an inference that is inappropriate to make (valid or not). Trying to effect a total information-preserving acceptance state transition in the face of transparently contradictory beliefs is generally unreasonable. The explanation for this can be simple: that in regulating one's beliefs, one should generally aim to have them be justified. In Ex Falso, one's starting attitudes are obviously untrue because contradictory. So engaging in deductive inference merely produces new beliefs having no basis in the truth, with no obvious countervailing epistemic virtues. This suffices to explain that the beliefs are unjustified, insofar as they are held merely on the basis of the contradiction. Giving this explanation, however, does not require saying anything about whether the inference is logically valid, or not, or a good inference (*qua* inference), or not. So we can give intuitive, negative answers to question (3) above about whether reasoning with Ex Falso is good, without making *any* commitments about whether Ex Falso is valid or is a good form of inference.

Ex Falso is valid, but rarely if ever a good form of reasoning. It's no surprise that these claims could be compatible, if the point of investigating validity is to investigate inference that is good *qua* inference, where that goodness is only sometimes appropriately exploited in good reasoning. But that leaves our second question: do, or can, inferences classified by Ex Falso count as good inferences? The answer on my view may be surprising. Recall that in addition to necessary-truth preservation, good inferences must be appreciated as necessarily truth-preserving to count as good. One might think that this appreciation is a minor added condition, since it generally only raises the question of whether a reasoner is in a position to cognitively grasp a relevant relation of necessary truth-preservation. But in the case of Ex Falso it intriguingly renders all inferences of the relevant form bad inferences, regardless of the cognitive sophistication of the reasoner who employs it. Indeed, the more sophisticated the reasoner, the less likely they will be in a position to make an inference of the relevant form at all.

On the account of inference I've offered, to infer q from $p \wedge \neg p$ in Ex Falso, would be to crowd-out a joint representation the contents $\neg q, p \wedge \neg p$ while both sensitive to the question of whether q and maintaining an accep-

tance of $p \wedge \neg p$. To be a genuine instance of Ex Falso, the relevant instance of crowding-out must be based in the 'form' of $p \wedge \neg p$ (as opposed to, say, being based on an appreciated lexical entailment from p to q). But then to be *correctly* performed, it must be based on crowding-out $p \wedge \neg p$, as it is only this impossibility (given that the inference is based on the relevant form) which can secure the necessary truth-preserving character of the inference. But then to perform the inference correctly is simply impossible. On the one hand, it requires maintaining acceptance of $p \wedge \neg p$ during the inference. And on the other, it requires crowding-out $p \wedge \neg p$ while one maintains this acceptance. But crowding-out $p \wedge \neg p$ precisely *precludes* entertaining that content, and so accepting it. So the inference cannot be correctly made.

Again, this shows the inference cannot be *correctly* made. Does this mean that one cannot ever infer as per Ex Falso? This is a slightly trickier question, whose answer will depend on what we count as an instance of inferring with the relevant 'form.' As discussed in Chapter 4, representational crowdings-out are not factive, and can involve a kind of mistaken perception of modal space. So this much may be possible: inferring q from $p \wedge \neg p$, solely because of the form of those statements (again, e.g., not because of a lexical entailment between p and q), but *without* appreciating that $p \wedge \neg p$ is impossible. One may, for example, 'take' $p \wedge \neg p$ to be possible in appreciation, but only at worlds where q happens to be true.

I am agnostic both as to whether this is a realistic kind of cognitive relation to get into, and as to whether transitions between such states should count as an instance of inferring by Ex Falso. The important point for now is this: there is no way of performing an inference with the form of Ex Falso while correctly grasping the entailment. Either one does not crowd-out relevant representations, or one crowds-out inadequately. In the first case, one actually does not infer, since inference constitutively involves crowding-out representations. Perhaps one does something else, like engaging in free association. In the second case, one infers badly, for failing to correctly crowd-out representations in the inference.

It follows that no inference with the form of Ex Falso is ever a good one. This shows just how far validity and good inference can come apart on the current view. It also reveals that each party to the debate on Ex Falso has a piece of the truth. Detractors are right to maintain it is a formal or technical excrescence. Ex Falso is a degenerate case arising from the need to provision-

ally set aside the psychological contingencies of appreciability. When we do this, we engage in logical inquiry to investigate only a necessary condition on good inference. But Ex Falso is a rare case in which this necessary condition on good inference can be satisfied, even though that condition can *never* be co-instantiated with further conditions sufficient for good inference. So there is a sense in which Ex Falso is indeed a 'mere artifact' of our methodological restriction. In spite of all this, defenders of Ex Falso are still perfectly within their rights to maintain the validity of inference. It is simply hopeless to try to avoid the methodological restriction which leads to the excrescence and instead capture the conditions on good inference more broadly (let alone the conditions on good reasoning) in a theory with the kind of rigor we see applied in formal logic. Good inference depends on a wide range of contingent psychological features that vary from agent to agent, and even time to time for a single agent. And good reasoning depends on all the complex epistemic norms governing belief formation generally (at least when the reasoning involves beliefs). Ex Falso is an excrescence, but an unavoidable one, at least if we want to maintain the kind of simplicity and rigor characteristic of logical inquiry while continuing to investigate inferential goodness. (Note: I am of course not claiming one cannot develop a simple enough *formal theory* in which Ex Falso fails. Relevantists have obviously done that. I am claiming that these formal theories as of yet have no well-defined and unified theoretical purpose *in the study of good deductive inference*—-or in the study of reasoning either, for that matter.) So there is no point in revising our theories to try to avoid the presence of Ex Falso. Doing so will simply warp the one legitimate and achievable theoretical purpose of logical inquiry in the study of inference.

The points I've been making so far are argued from within my particular conception of logic as the study of deductive inference, drawing on my particular conception of inference. So it may be worth pausing to get clear on where these arguments figure in the broader dialectic over the validity of Excluded Middle or Ex Falso. There are some obvious ways one could resist the views I've been arguing for. One is to follow me part way, acknowledging that logic is the study of inference, but defending a conception of inference which doesn't support the conclusions I've been drawing. Another is to get off the boat much earlier, and defend a conception of logic on which it does not have as its primary object of study patterns of good inference at all. The first dispute is a substantive one in the philosophy of mind. The second involves some mix

of mere terminological issues and more substantive issues, depending on how much of 'core' logical practice one wishes one's account of logic to be faithful to.

If one resists my argument along either dimension, there is an important methodological claim worth taking away. Even if I were forced to abandon either of my two key assumptions about logic, I would hold fast to the view that we should aim to resolve contested logical principles by first getting clear on what logic is, in broadly non-logical terms. This contrasts with the attempt to resolve contested principles by brute appeal to logical intuitions, which for obvious reasons makes little progress. But it also contrasts with recommendations like that of MacFarlane (and those he is responding to) to merely get clearer on the normative role of logic in order to settle such disputes. The virtue of MacFarlane's proposal is that it pushes questions away from intuitions into the foundations of logic. The problem with that recommendation is that it is in danger of only going half way.

Investigating norms on reasoning broadly construed, without saying anything more about the foundations of logic, only gets us to clear conclusions about logic on the presumption that logic has immediate and direct implications for reasoning broadly construed. Given my sympathies with some of Harman's points, it should be clear that I think that presumption is unlikely to be true. Still, it is at least an option to explore. But even if one takes this option, the foundational presuppositions needed to make an investigation of reasoning relevant to logic need to be stated explicitly and defended. Without that statement and defense, the use of norms of reasoning to settle logical matters is simply unjustified. Moreover, what the account here shows is that defending the relevance of norms of reasoning to logical matters is far from trivial. Indeed, the distance between logical pronouncements and reasoning broadly construed may be such as to make pronouncements on the latter essentially irrelevant to the former. This is true even if, as I have maintained, logic plays *some* highly important role in the study of reasoning. So, again, if we want to have investigations of logic be resolved by investigations of reasoning broadly construed, we should pair that view with some deeper foundational explanation of the connections between logical inquiry and reasoning. Failing that, we won't have any reason to think that looking only to norms of reasoning to settle questions of logic is sound methodology.

With that methodological point out of the way, let me take stock. The ex-

amples of Excluded Middle and Ex Falso aren't idiosyncratic. The characterization of logical consequence that I've offered gives us precisely the sort of purchase on foundational questions in logic that we should hope from an analysis or reduction: it transforms controversial questions about logic into questions in other non-logical domains, whose resolution doesn't immediately prejudge the original logical questions.

Of course, it is not obvious that *all* disputes about questions of logic will admit of this transformation. Here, something like Etchemendy's worry about circularity can resurface. For example, if we are fortunate, debates over the validity of the Law of Non-Contradiction may boil down to questions in the foundations of semantics as they do for questions about the Law of Excluded Middle. But if we are less fortunate, perhaps they will instead tend to reduce to the question of what kinds of metaphysical possibilities that there are—for example, whether there are metaphysically possible worlds where something is both in a particular spatial location and also not there. Perhaps the only thing to say in favor of the Law of Non-Contradiction, even at this foundational level in metaphysics, involves presupposing that law in a way that will be dialectically ineffective against its detractors.

This is a limitation of the utility of my characterization of consequence in arbitrating foundational disputes. And it does seem close to the kind of worries Etchemendy is expressing in the passage I quoted above. But I can't see how this limitation could in any way be billed a significant concern for the status of the characterization as something like a reduction. For example, the foregoing worry about the ineliminability of logical intuitions in discussions of the Law of Non-Contradiction would seem about as threatening to even the interpretationalist conception of validity, which Etchemendy praises for its reductive qualities. For some, the question of whether the Law of Non-Contradiction fails at the *actual* world is about as pressing as the question of whether it fails at merely *possible* worlds. Certainly some opponents of the principle do think it actually fails (notably Priest, owing to semantic paradox). If this is right, ground-level intuitions about logic will infect even interpretationalist construals of validity, with their focus on actuality.

What cases like this show is that we can't eliminate such ground-level appeals to logic at some point, for some disputes. That much should be obvious. If we were expecting a characterization of logic with reductive ambitions to avoid such appeals altogether, I can't help but think that we were transparently

expecting far too much.

The value of my reduction, as a reduction, lies in its ability to link logical questions to non-logical ones in a fruitful way. Accordingly, the view better meets Etchemendy's challenge to the extent it proliferates the number and nature of such links. The goal of the next chapters is in part to show that the benefits of the analysis aren't limited to a few select cases like Ex Falso and Excluded Middle. Rather, they extend to a whole host of logical issues arising in different frameworks, many of which concern not merely the justification of individual logical rules, but rather the interpretations or foundations of the logical frameworks taken as a whole. I'll begin this task in the next chapter, where we'll uncover some important ties between the interpretations of modal logics and views about the nature of linguistic content. These ties will help to draw out a logically-inspired puzzle about the conditions on good deductive inference in the philosophy of mind that will eventually bring logical questions into indirect contact with empirical questions about the semantics of attitude ascription.

CHAPTER 8

Validity in Modal Logics and A Puzzle about Inference

On the view I've offered, logical truths are metaphysically necessary. But Zalta (1988), drawing on Kaplan (1989b), has argued that logics with two-dimensional intensional operators reveal that some logical truths are metaphysically contingent.[1] What does my framework have to say about this issue?

After sketching Zalta's argument in §8.1, I note that its strength turns on the contested question of how to define validity in modal logics with two-dimensional operators. I then extend these considerations by further noting that Zalta's argument also interacts with the question of how the contents expressed by sentences of a modal logic relate to their compositional semantic values. Once both of these points are appreciated, I argue, the simplest version of Zalta's challenge can be resolved either by empirical considerations, or mere stipulation, and either way the metaphysical necessity of logical truths would be safeguarded.

The discussion reveals that there is an in-principle obstacle to arguing against the necessity of logical truth with broadly Zalta's strategy (at least on the view of logic I put forward). In particular, arguments based merely on the behavior of an *operator* cannot give us reason to abandon that claim. Still, the failure of these arguments also reveals some space to develop related arguments against the necessity of logical truth that focus on the nature of inference and the mental contents that inference operates on.

In §8.2, I develop a puzzle about inference related to the examples of Zalta, but that can be stated without reference to logic. Defending the claim that log-

[1] Some have attributed similar claims to Kaplan himself. I'll discuss this issue in much greater detail in Chapter 10, where we'll see that there is subtlety in Kaplan's commitments that complicates such an attribution.

ical truths are necessary against the strengthened puzzle requires me to take a controversial stance on the behavior of natural language attitude reports. This reveals an interesting way in which the application of logic, on the view I've developed, can be sensitive to empirical considerations.

8.1 Semantic Values, Assertoric Content, and Logics with Two-Dimensional Intensional Operators

Zalta (1988) notes that in modal logics containing certain rigidifying operators, such as a rigidified description operator or an actuality operator, we can find sentences that will be true on all Kripke-style interpretations,[2] even though the necessitations of those sentences can be false.

Consider a propositional logic with an actuality operator $\mathcal{A}$ and a necessitation operator $\Box$. Define an interpretation (or model) $\mathcal{I}$ for the language to be a tuple $\langle \mathcal{W}_\mathcal{I}, @_\mathcal{I}, V_\mathcal{I} \rangle$, where $\mathcal{W}_\mathcal{I}$ is a set of worlds, $@_\mathcal{I}$ a distinguished element of $\mathcal{W}_\mathcal{I}$ (intuitively: actuality), and $V_\mathcal{I}$ is a function from world, sentence letter pairs to (bivalent) truth-values (intuitively: a truth-value assignment to propositions expressed by sentence letters at each world). In what follows, I will often suppress the interpretation subscript.

We recursively extend the valuation V over sentence letters to a valuation V' over all sentences using the natural clauses for connectives plus the following two for our operators: $\Box\phi$ is true at w iff ϕ is true at all $w' \in \mathcal{W}$ (thus building the S5 interpretation of the necessity modal into the semantics for simplicity); $\mathcal{A}\phi$ is true at w iff ϕ is true at $@$.

Say a sentence S is true on an interpretation at a world w just in case the interpretation's extended valuation V' maps S and w to truth. Say a sentence S is true on an interpretation just in case S is true at $@$ on $\mathcal{I}$. And say a sentence is *real world valid* just in case it is true on all interpretations.[3]

Then consider the following sentence, interpreting the sentence letter p as a contingent fact, with accompanying intuitive paraphrase.

(1) $p \rightarrow \mathcal{A}p$

(1′) If Biden is visiting China, Biden is visiting China at the actual world.[4]

[2] Kripke (1963).

[3] As recently noted, the characterization of validity is broadly Kripkean though, for reasons that will become clearer soon, the terminology follows Davies & Humberstone (1980).

[4] I use the stilted formulation "at the actual world" for relativizations to actuality, rather

(1) is real world valid. For any interpretation, $p \rightarrow \mathcal{A}p$ is true at @ just in case either p is false at @, or $\mathcal{A}p$ is true at @. But $\mathcal{A}p$ is true at @ just in case p is true at @. And p must be either true or false at @, given our assumption of bivalence.

By contrast (2) can be false on some interpretations.

(2) $\Box(p \rightarrow \mathcal{A}p)$

(2′) Necessarily, if Biden is visiting China, Biden is visiting China at the actual world.

To show this, it suffices to pick any interpretation on which p is false at @, but true at some $w' \neq @$. At w', $p \rightarrow \mathcal{A}p$ will be then by false, since p is true at w', but false at @. And that is enough to falsify $\Box(p \rightarrow \mathcal{A}p)$ at @.

Thus there are sentences of our formal system that are real world valid, whose necessitations are false on some interpretation. *If* real world validity is a form of 'genuine' logical validity, and *if* false necessitations reveal their complements to be contingent, then we will have sentences that are logically valid but contingent. But do these two conditions hold?

Zalta is duly cautious, and recognizes the need for argument here. Indeed, the need for argument is pressing since there is a rival formalization of validity for modal logics sometimes termed *general validity*.[5] A sentence is generally valid just in case it is true on all interpretations at *all* worlds. It is easy to see that sentences which are validities in this sense have valid necessitations: $\Box\phi$ will be generally valid just in case it is true on any interpretation at any world. That will hold just in case ϕ is true on any interpretation at any world—which is precisely what it takes for ϕ itself to be generally valid.

Zalta gives two reasons for preferring real world validity over general validity.

> (1) the most important semantic definition for a language is the definition of truth under an interpretation, and the [method of characterizing general validity], in which no world is distinguished as the actual world, has no means of defining this notion; and (2) the semantic notion of logical truth is properly defined in

than "actually", in trying to tease out informal intuitions, since there is evidence that the English "actually" has an established meaning on which it never receives the interpretation of our desired operator—see Yalcin (2015).

[5] Again following Davies & Humberstone (1980).

> terms of the semantic notion of truth, and the alternative definition of logical truth [i.e., general validity] is the wrong one because it fails to do this.
>
> (Zalta, 1988, 11)

Zalta claims it was a key Tarskian insight that logical truth is truth on all interpretations (full stop). We can define a notion of truth on an interpretation (full stop) using the privileged status of truth-relative-to-the-actual-world-on-an-interpretation. But general validity is defined not in terms of truth on all interpretations, but truth on all interpretations *at all worlds*. Zalta notes that some logicians define models without privileging an actual world, and define logical truth as truth-in-all-models-at-all-worlds without an intermediate definition of truth on an interpretation (full stop). And this, according to Zalta, relinquishes an important insight of Tarski's.

> Tarski's insight is that logical truth is truth-under-all-interpretations. When we define a modal language, we need a definition of truth-under-an-interpretation if we are to define logical truth by applying Tarski's insight. Kripke's models permit the definition of logical truth to follow this pattern (the alternative models [i.e., those used in defining general validity] do not)…So we do not beg any questions when we argue that the alternative definition of logical truth [i.e., general validity] is incorrect. It fails to take the notion of truth seriously.
>
> (Zalta, 1988, 11)

Following Hanson (2006), I find this argument unpersuasive.

As Hanson notes, we can easily get the desired properties of general validity using a notion of truth on an interpretation (full stop), as long as we shift to interpretations with *two* distinguished worlds: @ and w^d. These worlds each take over half of the dual role played by the actual world in the characterization of real world validity: that of giving the semantics for the actuality operator, and that of being the world relative to which truth-on-an-interpretation is settled. On our new interpretations, @ continues to be used in the recursive clauses of the actuality operator, but it is w^d that is used to evaluate the truth of a sentence on an interpretation. A sentence of our simple modal logic with an actuality operator will be generally valid just in case is 'true on all interpretations' (full stop), where an interpretation has these two distinguished worlds,

and the second distinguished world anchors the 'unrelativized' definition of truth on and interpretation.

Hanson, rightly, does not rest his resistance to Zalta merely on this point, and notes the obvious line of resistance.

> Zalta would probably respond that [the redefined notion of general validity] does not meet his Tarskian requirement for a definition of logical truth because it is important that the designated world of an interpretation, the world in terms of which truth under that interpretation is defined, always be the actual world of that interpretation.
>
> (Hanson, 2006, 443)

Hanson frames this reply on behalf of Zalta in terms of the importance of how we construe an 'interpretation.'[6] But it would be fruitless to argue that we should restrict our attention to truth at the actual world merely on the basis of how the technical word "interpretation" is used. That word does not have an established pre-theoretical use that could bear much argumentative weight. Rather, the concern is more revealingly put in terms of how interpretations are used to track validity: that the kind of truth, relative to which logical truth is assessed, should always be truth at actuality. This does seem consonant with Zalta's emphasis on unrelativized truth-on-an-interpretation, since actuality is the implicit world relative to which unrelativized truth is typically assessed in first-order theorizing.

We should distinguish the claim here that actual truth is integral to the characterization of logical validity from a similar claim we saw in Chapter 7 operative in the so-called 'interpretationalist' construal of first-order model theory. There we saw how something like the following characterization was integral to the interpretationalist stance.

> A sentence S is logically true iff S is actually true, no matter how the non-logical vocabulary in S is interpreted.

This, of course, cannot be the construal of logical truth that Zalta has in mind, at least if the 'actual world' in his logical apparatus models anything like the

[6] The ensuing response here differs slightly from that Hanson gives, but is certainly similar in spirit.

actual metaphysically possible world we inhabit. The reason is that in determining even real world validity, we allow the value of the actual world to *vary*. Thus in determining logical truth, we assess sentences' semantic properties at many clearly counterfactual circumstances. It's just that these are treated as 'actual' as we vary interpretations. Once we see this, it's clear we can't retain both an interpretationalist construal of validity generally, and anything like the standard semantic characterizations of validity for modal logic (a point Etchemendy emphasized in building his case against interpretationalist construals of logic generally).

What Zalta must lean on is not the importance of the notion of 'truth at the actual world,' but instead that of truth at a (possibly counterfactual) world *considered as actual*. A bit metaphorically, when we ask about the truth of some sentence relative to a possible world w, we consider some sentence and ask whether (given certain aspects of its current semantics) it would come out true were it evaluated 'from within' that particular world w, so that (among other things) w becomes the anchor for the actuality operator. The notion here goes back to what Evans (1979) terms "truth in a possible situation" and contrasts with what Evans calls "truth with respect to a possible situation". The latter allows a sentence in a world to be evaluated for truth relative to a possibly different choice of world as actual.

Once we see that Zalta must lean on Evans's notion of truth-in-a-world, though, the force of his rhetoric vanishes. Zalta claims that he is merely extending a 'key insight' of Tarski to the modal case. But there is no way in which the standard Tarskian characterizations of validity support the notion of truth-in-a-world to the exclusion of that of truth-with-respect-to-a-world. In Chapter 7, we noted that there are two broad construals of those Tarskian characterizations: the interpretationalist construal and the representationalist one. But neither construal supports Zalta's claim. On the one hand, as we've just seen, the interpretationalist construals of first-order machinery are far too stringent to extend to the modal setting while respecting anything like either real world or general validity. And on the other hand, representationalist construals of first-order interpretations are constitutively grounded in a *relativized* notion of truth: truth relative to a possibility. Those possibilities need not be metaphysical possibilities, as I've maintained. The important thing is that on a representationalist view, logical truth is grounded in a relativized notion of truth despite surface appearances: varying 'actual' interpretations (and so con-

sidering only truth on an interpretation, full stop) in the first-order setting is merely a convenient way of modeling the relativized notion of truth. And what is more, on the representationalist construal, as we vary an interpretation to model truth relative to a new possibility, the actual world plays no distinctive role.

This is not yet to argue *against* the notion of real world validity. It is only to note that Zalta's arguments for it fall flat. It is correct that truth full stop, or actual truth, plays a distinguished role in standard first-order interpretation. But the role it plays is either too stringent to play in modal logics, or a merely instrumental role in modeling a relativized notion of truth. Either way, the privileging of actual truth in first-order interpretation cannot give any distinctive support for real world validity.

This still leaves the danger, for my view, that some other argument for real world validity could be devised. So can we bolster our case by arguing against real world validity? Hanson (2006, 445–7) tries to do so, attempting to use Zalta's own commitments against him. Hanson, like Zalta, takes logical truths to be analytic, where analyticity should amount to truth 'merely' in virtue of meaning. But, Hanson claims, real world validities are not true merely in virtue of their meanings, but rather true in virtue of their meanings relative to a selection of a possible world as the actual one. General validity, by contrast, captures truth in virtue of meaning *simpliciter*.

I will not here take a stand on whether logical truths are analytic, in part because it is not necessary to evaluate Hanson's argument. Even if logical truths *are* analytic, this would not tell against real world validity. Here is Hanson's argument (paralleling one made by Zalta against Kaplan) that real world validity fails to capture analytic truths:

> We cannot consider the truth of the sentence $[p \rightarrow \mathcal{A}p]$ without appealing to some actual-world candidate, and so we cannot simply say that it has the property that traditional analytic truths have, namely, being true in virtue of the meanings of its words. Rather, it has the property of being true in all actual-world candidates in virtue of the meanings of its words relative to such actual-world candidates.
>
> (Hanson, 2006, 446)

The "and so" in this passage does not seem to me to be justified. It would seem that a sentence whose truth can only be assessed relative to something (a time,

a world, the selection of an actual world, an interpretation of the non-logical terms) could still be true in virtue of its meaning, if any logical truths have that property. What would contribute to such a sentence's being true merely in virtue of its meaning would be for the value of the parameter in question *not to matter*—for the sentence to be true regardless of how we select the value of the parameter relative to which truth is ascertained.[7] But that is precisely what real world validity is: truth secured regardless of how the actual world is selected. Indeed, when we evaluate $p \rightarrow \mathcal{A}p$ in considering Hanson's preferred general validity, we also must assess it for truth relative to the selection of an actual world. It's just that we must *in addition* assess it relative to a choice of a possible world of evaluation that may differ from the selection of the actual world. How does the double relativization make the problem go away?

To press this point further: assuming, as we are in this setting, that sentences have truth-values relative to a possible world, I fail to see how Hanson's argument fares any better than the one below, which surely shows too much.

> We cannot consider the truth of the sentence $p \rightarrow \Diamond p$ without appealing to some *possible world candidate*, and so we cannot simply say that it has the property that traditional analytic truths have, namely, being true in virtue of the meanings of its words. Rather, it has the property of being true in all *possible world candidates* in virtue of the meanings of its words relative to such *possible world candidates*.

We should not want to give up the analyticity of $p \rightarrow \Diamond p$. At the very least Hanson is committed to this. But we seem to have parity of reasoning, as far as I can see. So I think that Hanson's argument for general validity, like Zalta's for real world validity, fails on its own terms, this time for involving a simple non-sequitur. Indeed, if anything there is a worry that Hanson's argument backfires. Once we reflect on the argument's failure, we can see that real world validity may indeed capture (mere) truth in virtue of meaning in the proposed system, so that $p \rightarrow \mathcal{A}p$ is analytic, at least on the construal Hanson himself seems to prefer.[8] It is a sentence which expresses a truth merely owing to its

[7] Cf. the defense of the notion of truth-in-virtue-of-meaning in Russell (2008) Ch.1.

[8] Well: truth in virtue of meaning *plus* the other customarily ignored syntactic and semantic assumptions being built into the relevant interpretations, such as those I discussed under the heading of modalized first-order form in Chapter 7.

meaning, with no substantial contribution from how the world is: no matter which world we consider the sentence at, what it expresses is a truth.

Since both Zalta's and Hanson's arguments fail, we are at an impasse. So we continue to be left with the question: if logic is construed in my terms, which better captures logical validity in the posited modal framework—real world or general validity?

Given the apparently high stakes raised by Zalta's challenge, the answer is perhaps surprising. My framework on its own favors *neither* of these two conception of validity. It is compatible with either. What is more, regardless of which conception of validity is adopted, there is no immediate threat to the claim that logical truths are necessary.

To understand why all this holds, we need to revisit some ideas from Chapter 7. A key lesson of that chapter was that we cannot ask, let alone answer, questions about validity in a logical framework until we have settled what truth-conditional assertoric contents are expressed by the sentences of that framework. In the case of first-order logic, the need to settle this issue was transparent. Sentences in first-order models are only given extension-level semantic properties, including single truth-values. There is simply not enough information in a first-order model to recover truth-conditions relative to metaphysical possibilities. Accordingly, we had to make some choices about what those truth-conditions would look like by trying to generalize the semantic properties characteristic of first-order models to a world-relative setting. This led us to the notion of a modalized interpretation. Only once we fixed the expression of truth-conditions using such interpretations could questions about validity, in my sense, have non-trivial answers.

In the case of broadly Kripkean interpretations for modal logics, unlike in the first-order case, it may *seem* that these issues about truth-conditional content are already settled by the interpretations themselves. After all, the interpretations for modal logics relativize semantic properties of expressions to a world. As long as the 'worlds' of the interpretation can correspond to metaphysical possibilities, then checking for truth in all interpretations may be giving us information about metaphysical necessity.

But a moment's reflection reveals many obvious complications with construing the worlds of interpretations as metaphysical possibilities. For example, in assessing validity in a modal logic, we are allowed to vary the number of possible worlds in an interpretation. But it is not unreasonable to think

that the number of metaphysically possible worlds is fixed. Moreover, an interpretation may attribute to a set of worlds varying degrees of combinatorial complexity. But how much combinatorial variation there is within the space of genuinely metaphysically possible worlds would also appear to be a fixed matter. Worse, it is controversial just how much, and what kind of, combinatorial variation metaphysical modal space exhibits. For example, Lewis (1986) claims that metaphysically possible worlds satisfy a combinatorial condition of *plenitude*: any recombination of elements from a possible world should itself correspond to a possibility. Other philosophers, like Armstrong (1989, 1997), advance rival permutational combinatorial principles, in which worlds should exhibit sufficient variability in how fundamental properties are instantiated.[9] This is not to mention the complex question of what kind of accessibility relation among worlds should be built into the modality to properly reflect 'genuine' metaphysical modal space.

These concerns do not show that Kripkean interpretations are unhelpful in tracking metaphysically necessary truth in virtue of certain linguistic properties. Rather, they show that the question of whether they do is non-trivial, and settling the answer requires argument. This is true of the modal setting just as much as it was of the first-order setting. In particular, we continue to need some form of argument that explains how variation among the models either corresponds to genuine variation within the space of metaphysical possibility or variation in linguistic properties 'ignored' for logical purposes. Or, if we cannot provide such an argument, we need an argument to explain why the variation in interpretations can, by some other means, continue to track metaphysically necessity in virtue of some set of linguistic properties, just as we saw was possible in the first-order case.

I want to set these issues aside for the moment. This is because, for now, there is an *additional* respect in which the use of model-theoretic interpretations must be treated with care in defining validity. This is that a model-theoretic interpretation is most clearly giving us a recursive characterization of truth relative to shiftable parameter—in this case a world parameter. That is, we are getting information about what are sometimes called the 'compositional semantic values' of expressions in our language with modal operators. What

[9] For further discussion of recombination principles see, e.g., Forrest & Armstrong (1984), Nolan (1996), Sider (2005), Efird & Stoneham (2008), Wang (2013), Russell & Hawthorne (2018).

we need to start assessing claims about validity, I've claimed, is information about assertoric contents. But there is a gap here between the compositional semantic values supplied in the model-theoretic interpretations and the assertoric content that is our true object of concern.

It is a familiar point in the philosophy of language that we need to be careful to distinguish compositional semantic values from assertoric content.[10] It is easy to conflate these two notions, because the compositional semantic value of a whole sentence is obviously closely related to the assertoric content it expresses. In particular, the semantic value should *determine* that content (perhaps alongside other contributions, say, from context). But there are often different, incompatible ways in which a compositional semantic value can be used to determine assertoric content. This point is especially true when we move to modal logics that contain 'two-dimensional' operators, like an actuality operator (though we should bear in mind the point even holds for Kripkean interpretations of modal logics that don't contain such operators).[11]

Note that on any interpretation given for a language with an actuality operator, we will assess the truth of sentences of the form $\mathcal{A}p$ with respect to two worlds: a world 'of evaluation' and a world designated as actual. This dual relativity of the semantics of the actuality operator familiarly yields two different ways of using a the two-dimensional framework to yield truth-conditional content. On the one hand, we can anchor the actual world while taking truth-conditional content to be specified by variation in the other world parameter. Alternatively, we can allow both the actual world and the other world of evaluation to covary while settling truth-conditional content.

To be a little more precise, let $\mathcal{I}_w$, for w in $\mathcal{W}_\mathcal{I}$, be the interpretation differing from $\mathcal{I}$ at most in that w is assigned to the role of the privileged 'actual' world. Then, considering sets of worlds as truth-conditions, we have at least these two candidates to play the role of the assertoric content expressed by a sentence ϕ at a world w (relative to interpretation $\mathcal{I}$).

The Horizontal View: The truth-conditional content expressed by ϕ at

[10] See Dummett (1959), Lewis (1980), Stanley (1997), Ninan (2010b), Rabern (2012), and Yalcin (2014).

[11] A two-dimensional semantics is one on which there are at least two (noteworthy) variable determinants of an expression's extension. A two-dimensional operator an operator which is responsive to both of those determinants.

$w \in \mathcal{W}_\mathcal{I}$ is

$$[\![\phi]\!]^{\mathcal{I},w}_H = \{w' \in \mathcal{W}_\mathcal{I} \mid V'_{\mathcal{I}_w}(\phi, w') = t\}$$

The Diagonal View: The truth-conditional content expressed by ϕ at $w \in \mathcal{W}_\mathcal{I}$ is

$$[\![\phi]\!]^{\mathcal{I}}_D = \{w' \in \mathcal{W}_\mathcal{I} \mid V'_{\mathcal{I}_{w'}}(\phi, w') = t\}$$

On the Horizontal View of content, a privileged world w plays the 'anchoring' role I mentioned before. On this view, the sentence expresses different contents at different worlds, considered as actual. On the Diagonal View, the actual world no longer anchors content, but covaries with the non-designated worlds in a set of interpretations to determine a set of truth-conditions. Note that on the Diagonal View, the sentence ϕ expresses the same truth conditional content relative to every world in $\mathcal{W}_\mathcal{I}$. The broad motivations for these two conceptions of content are familiar from discussions of two-dimensional semantics, though the construal of the two-dimensional frameworks can vary in important respects.[12]

It is now natural to ask: which is the correct view of the assertoric content of a sentence of our modal logic? The answer will depend on our theoretical aims. We could take our formal language to be modeling some aspects of natural language use. If so, then the question of what conception of assertoric content is 'correct' may be sensitive to empirical linguistic concerns, such as how natural language users employ the expressions we are modeling. It may equally be sensitive to foundational questions about the nature of assertion.

I don't want to speak to any of these issues yet. Instead, I want to note that it seems conceptually possible to use sentences to assert either their horizontal or diagonal content.[13] If this is right, then as soon as we leave empirical considerations behind, we are free to simply *stipulate* either conception of assertoric content as part of the intended construal of the logical machinery, not unlike

[12] For example, two-dimensional frameworks can be interpreted along contextual (Kaplan, 1989b), metalinguistic (Stalnaker, 1978), or epistemic (Chalmers, 2004) lines.

[13] I hedge with "seems" because one could try to make the case that the very nature of assertion favors one of these views so that any practice which used sentences to express the alternative kind of content would somehow necessarily fail, or at least fail to count as a practice in which assertions were made.

the way we might stipulate the semantic behavior of an operator within the language.

With such a stipulation, we would at last be in a position to ask and address questions about validity. And it is not hard to see that opting for one or the other of the proposed views of assertoric contents corresponds quite directly to our competing conceptions of validity for the language.

For example, the horizontal conception of assertoric content vindicates something like general validity and consequence. Given an interpretation $\mathcal{I}$, the set of worlds that is the horizontal truth-conditional content expressed by p relative to a world is not always a subset of the set of worlds comprising the horizontal truth-conditional content of $\mathcal{A}p$. For example, the horizontal content expressed by $\mathcal{A}p$ at a world w will be the empty set provided the content expressed by p at w is false at w. But the horizontal content expressed by p at w could yet be true at some worlds. And it is not hard to see that for this very reason the inference from p to $\mathcal{A}p$ is blocked on the general construal of validity. By contrast, the set of worlds comprising the diagonal truth-conditional content of $\mathcal{A}p$ is always simply *identical* to the set of worlds comprising the diagonal truth-conditional content of p. And it is not hard to see that this is the very reason that the inference from p to $\mathcal{A}p$ is permitted on the real world conception of consequence.

More generally: ψ is a general consequence of ϕ just in case for all interpretations $\mathcal{I}$, the transition from the horizontal content of ϕ at w on $\mathcal{I}$ to that of ψ at w on $\mathcal{I}$ preserves truth at all worlds in $\mathcal{W}_\mathcal{I}$; and ψ is a real world consequence of ϕ just in case for all interpretations $\mathcal{I}$, the transition from the diagonal content of ϕ at w on $\mathcal{I}$ to that of ψ at w on $\mathcal{I}$ preserves truth at all worlds in $\mathcal{W}_\mathcal{I}$. This essentially follows by definition.

ψ is a general consequence of $\phi \Leftrightarrow$
$\forall \mathcal{I} : \{w' \in \mathcal{W}_\mathcal{I} \mid V'_\mathcal{I}(\phi, w') = t\} \subseteq \{w' \in \mathcal{W}_\mathcal{I} \mid V'_\mathcal{I}(\psi, w') = t\} \Leftrightarrow$
$\forall \mathcal{I}, \forall w \in \mathcal{W}_\mathcal{I} : \{w' \in \mathcal{W}_\mathcal{I} \mid V'_{\mathcal{I}_w}(\phi, w') = t\} \subseteq \{w' \in \mathcal{W}_\mathcal{I} \mid V'_{\mathcal{I}_w}(\psi, w') = t\} \Leftrightarrow$
$\forall \mathcal{I}, \forall w \in \mathcal{W}_\mathcal{I} : [\![\phi]\!]_H^{\mathcal{I},w} \subseteq [\![\psi]\!]_H^{\mathcal{I},w}$

ψ is a real-world consequence of $\phi \Leftrightarrow$
$\forall \mathcal{I}$: if $V'_\mathcal{I}(\phi, @_\mathcal{I}) = t$, then $V'_\mathcal{I}(\psi, @_\mathcal{I}) = t \Leftrightarrow$
$\forall \mathcal{I}, \forall w' \in \mathcal{W}_\mathcal{I}$: if $V'_{\mathcal{I}_{w'}}(\phi, w') = t$, then $V'_{\mathcal{I}_{w'}}(\psi, w') = t \Leftrightarrow$
$\forall \mathcal{I} : \{w' \in \mathcal{W}_\mathcal{I} \mid V'_{\mathcal{I}_{w'}}(\phi, w') = t\} \subseteq \{w' \in \mathcal{W}_\mathcal{I} \mid V'_{\mathcal{I}_{w'}}(\psi, w') = t\} \Leftrightarrow$
$\forall \mathcal{I} : [\![\phi]\!]_D^\mathcal{I} \subseteq [\![\psi]\!]_D^\mathcal{I}$

Note that a transition from ϕ to ψ holding fixed a *particular* interpretation $\mathcal{I}$

may preserve truth on either the horizontal or diagonal conceptions without the inference being licensed by general or real world consequence respectively. (Just consider an interpretation $\mathcal{I}$ whose valuation V maps two sentence letters p and q to truth in all worlds in $\mathcal{W}_{\mathcal{I}}$.) This intuitively corresponds to the fact that the necessary condition on good inference that logic helps track may be secured by meanings other than those belonging to merely logical vocabulary.

What this is showing us is that on my construal of logic, it turns out that general and real world validity are not necessarily *competing* conceptions of validity, at least if key empirical questions are bracketed. Rather, they are both acceptable conceptions relative to substantive choices about the construals of the assertoric content of the sentences of our modal logic. And *nothing about the standard semantics for the logic itself settles that substantive issue*.

What happened to Zalta's worry, seemingly acknowledged by Hanson, that if real world validity prevailed we would have contingent logical truths? This worry has largely evaporated, as soon as we drew the distinction between compositional semantic values and assertoric content. If we adopt the real world conception of validity $p \rightarrow \mathcal{A}p$ will indeed express a validity, while $\Box(p \rightarrow \mathcal{A}p)$ may well be false on some interpretations. But this only holds because the diagonal assertoric truth-conditional content expressed by the first sentence—content that is true at all metaphysical possibilities—is not what is being evaluated for necessity by the modal operator $\Box$. Rather, the modal operator is checking for the necessity of something that resembles the (would-be) horizontal content of that sentence (bearing in mind that this is a loose gloss: we needn't think of the operator as operating on 'content' at all, and it is perhaps best to avoid that construal). That is, the necessity modal keeps the actual world fixed, while evaluating $p \rightarrow \mathcal{A}p$ relative to counterfactual circumstances. We might accordingly label this operator $\Box_H$. The behavior of $\Box_H$ ensures that it fails to track information about assertoric content, at least given our stipulation of that content as diagonal. So there is no conflict between saying that logical truths are necessary, that $p \rightarrow \mathcal{A}p$ is a logical truth, and that $\Box_H(p \rightarrow \mathcal{A}p)$ is false. The claim that logical truths are necessary is a claim about assertoric content which is not contravened by the truth of $\neg\Box_H(p \rightarrow \mathcal{A}p)$—indeed the latter now says nothing bearing on the properties of assertoric content at all.

There is as yet no 'language-internal' operator which tracks aspects of assertoric content, given the assumption that sentences have their diagonal asser-

toric contents. But just as much as we can stipulate what form of content—diagonal or horizontal—is expressed by sentences of our logic (again bracketing empirical questions), we can stipulate how our modal operators behave. If we like, we can introduce a new operator $\Box_D$ which intuitively checks for the necessity of its complement's diagonal content. This operator, unlike $\Box_H$, would shift which world anchors the evaluation of $\mathcal{A}$ as it checks for truth at all possible worlds. Indeed, Davies & Humberstone (1980) define an operator that does precisely that. Any real world validity ϕ will be such that $\Box_D\phi$ is always true, and indeed valid.

What this means is that, so long as we bracket empirical considerations, the appearance of contingent logical truths in modal systems is, on my conception of logic, an illusion created by a mere mismatch between assertoric content and the behavior of language-internal modal operators. Restoring a match between them reinstates the necessity of logical truth *no matter how* we construe assertoric content. Once we appreciate this point, we can see that both Zalta and Hanson are mistaken to think that the necessity of logical truth turns on the choice of real world or general validity. That choice is so far simply irrelevant.

I should stress again that all this holds *if* we bracket empirical considerations. There may be an important worry in the vicinity of the considerations raised by Zalta that requires more than mere distinction-drawing once those empirical considerations are reintroduced. In §8.2, I will try to draw that problem out. But it is critical, before examining that issue, to get clear about some ways in which problems cannot arise. Otherwise, we could become embroiled in fruitless disputes.

Questions about the necessity of logical truths, on my view, are questions about assertoric contents insofar as they can figure as mental contents. So it is only at the level of content that the threat of logical contingency could arise. Let's turn now to see how just such a threat could surface by shifting gears to connect Zalta's examples to some important issues in the philosophies of mind and language.

8.2 A Puzzle about Inference

If the behavior of rigidifying operators, like an actuality operator, are to apply pressure to the claim that logical truths are necessary, it seems unhelpful to try

to show this by appeal to their interaction with a language-internal necessity predicate. To begin, any characterization of validity for a modal logic should track a corresponding conception of assertoric content. Once those conceptions of validity and content are fixed, the behavior of a language-internal necessity predicate is only relevant to the necessity of logical validities insofar as it tracks properties of their assertoric content. That connection would need to be argued, before the behavior of the operator could have any relevance to validity. But arguing for such a connection would certainly reveal that any objections we could raise to the claim that logical truths are necessary using the operator could have also been raised merely at the level of content, by showing that the content itself is not true relative to all possible worlds. So consideration of the language-internal necessity predicate would be superfluous.

Of course, what this leaves open is that there may be a way of using rigidifying operators to argue against the view that logical truths are necessary by focusing more directly on the contents involved in inference. My goal in this section is to formulate just such an argument, and develop some resources to defend against it. It turns out that the argument I consider forms part of a larger puzzle that creates a tension between two plausible claims: one in the philosophy of mind about the nature of deductive inference, and another in the semantics of attitude reports and their linguistic complements. To defend my account of inference, I must reject the claim about attitude reports. Whether one follows me in making this rejection, the puzzle reveals some ways in which thorny empirical issues bear on the study of inference, and so on the nature of logic as I conceive of it.

If we want to explore how rigidified operators can teach us anything about inference, we have to investigate how their behavior can be revelatory of properties of mental content. As I noted in Chapter 2, logic studies inference under the presumption that the contents that attitudes take as objects can also be expressed by sentences of a language. So the next step is to consider whether inferences which are reported using sentences containing rigidified operators can apply pressure to my claim that inference aims at a necessary truth-preserving transition of acceptance states. For these purposes, I'll make use of rigidified descriptions, rather than a sentential actuality operator, since I think the former can bring out intuitions a little more clearly.

Consider a reasoner, Belle. Belle starts out by believing two things, expressed by (1) and (2).

(1) Trump is the person who is president of the US in 2022.

(2) Trump is in Beijing.

Belle then infers a new belief from these two: the belief in what is expressed by (3).

(3) The person who is president of the US in 2022 in the actual world is in Beijing.

This is apparently all expressible in English as follows:

(B) Belle believes that Trump is the person who is president of the US in 2022 and believes Trump is in Beijing. She infers from these beliefs that the person who is president of the US in 2022 in the actual world is in Beijing.

Let's set aside questions in logic—in particular about validity—and only ask a question about the inference reported in (B). As I've stressed in previous chapters, we should distinguish the question of whether an inference is good *qua* inference from the question of whether it was a good idea to perform the inference and whether it shows a reasoner to have reasoned well. We know Belle's first belief is false. Perhaps Belle's first belief is also completely and transparently unjustified given her evidence. If so, Belle is not reasoning in the way she should, and is performing inferences she ought not to. But even if so, we can still ask, *given* that she is inferring: is she doing it correctly?

I submit that, considered on its own, Belle's inference is unproblematic: it is a safe way for Belle to expand the information in her initial store of beliefs in an appreciable and reliable way. We can suppose Belle has the cognitive resources to appreciate the goodness of her inference, and exploits them. And it isn't hard to see that some logics will predict Belle's inference is a good one. As we've seen, natural formalizations of Belle's reported beliefs in a modal logic would fall under the relation of real world consequence.

I noted in §8.1 that adopting real world consequence as our form of validity didn't, on its own, threaten the claim that logical consequence should be necessarily truth-preserving. For claiming that real world validity is a form of true validity is compatible with claiming that consequence necessarily preserves truth provided that the contents logic characterizes are diagonal contents. As long as we weren't investigating any particular natural language, I claimed, we were

seemingly free to stipulate whatever conception of content we liked, and there would be an associated conception of validity to track it.

But now we are considering contents *actually* reported in a natural language—English. So if we are developing a logic on the basis of that language, we must do justice to the actual contents expressed by its sentences, both in assertions of sentences like (1)–(3), and in characterizing any content that is reported by the complements of attitude verbs as in (B).[14]

Here is the worry. It is a common view—perhaps the default view—that assertions in English express something like what I've called their horizontal truth-conditional content.[15,16] On this view, (3) as asserted at the actual world would express a proposition that is true at a metaphysically possible world w just in case Joe Biden is in Beijing in w.

But if (B) reports Belle as believing such horizontal contents in the course of her inference, that inference will not be truth-preserving at all metaphysical possibilities. (1) and (2) would jointly be true at all worlds where Donald Trump is the president of the US in 2022 while being in Beijing in those worlds. (3) would be true at worlds where Biden is in Beijing. But there are metaphysically possible worlds where Trump became president and is in Beijing while Biden is not in Beijing.

What this tells us is that *if* (B) reports Belle as believing something like the horizontal contents of the relevant attitude-verb complements, and *if* (as it appears) Belle's deductive inference is a good one, then good deductive inference needn't be necessarily truth-preserving. Indeed, we can come up with many other inferences of roughly the 'form' of (B) (varying their 'non-logical vocabulary') that appear just as good as Belle's. So this appears to be just the kind of inference we might like to capture with logical frameworks like those discussed in §8.1.

[14] I am speaking of attitude verb complements as having content, which is a bit sloppy. It might be more precise, e.g., to talk instead about "the content such-and-such attitudes must have as objects if an attitude report, in which verbal complement has such-and-such a semantic value, is to express a truth." But looser talk should be fine here for the sake of simplicity.

[15] I am glossing over at least two major complications here. First, the compositional semantics of all the sentences I am considering are contested. Second, there are many different conceptions of diagonal content, and possibly horizontal content as well (see n.12). Still, I do not think these complexities substantially influence the problems I am about to describe, or the rough forms their solution could take.

[16] Note, I am not presuming that assertoric contents *are* truth-conditions. Maybe they are structured contents as some Fregeans and Russellians would maintain. I am only presuming that assertoric contents *have* truth-conditions.

This finally begins to apply some pressure to the view of inference I've developed in Part I. To maintain that view, one must either deny the goodness of Belle's inference, which is highly counterintuitive, or one must make a noteworthy commitment about the kinds of content reported by English sentences when they figure as complements in attitude reports, which will be highly controversial.

Still, I think one of those two claims—the claim that Belle's inference is no good, or the claim that (B) does not report Belle as believing the horizontal contents of the relevant attitude verb complements—must be the correct. What I would like to do in the remainder of this section is to bring out the case for the disjunction, and begin to reveal some of the complexity in the linguistic matters that bear upon this issue.

To begin, let me make the disjunction a little more specific. If Belle's inference is in fact good, I must deny that (B) reports Belle as believing the horizontal contents of the relevant attitude verb complements. How? It should be obvious, in light of the work of §8.1: by embracing the view that in attitude reports like (B), something like the diagonal contents of the verb complements are reported as believed. For convenience, let's call a view of broadly this form *the doxastic diagonal view*.[17] So the view I want to explore is expressed in the following disjunction: either Belle's inference is bad or the doxastic diagonal view is true.

Why think this disjunction holds? The first thing to note is that Belle's inference is actually one half of a pair of inferences which jointly raise a puzzle. Consider Suparna. Suparna has a capacity to infer—and in particular a capacity for appreciability—identical to that of Belle. Moreover, Suparna performs an inference that looks superficially similar to that of Belle, but for one key difference: all Suparna's attitudes come in the form of counterfactual suppositions. She counterfactually supposes the contents expressed by (1) and (2). She then deductively infers, under counterfactual supposition, the content expressed by (3). Let me suppose for now that this would be expressed in English as follows.

(S) Suparna supposes that Donald Trump were the person who is president

[17] Note: the doxastic diagonal view takes no stance on what is asserted with (1)–(3). A tempting extension of the doxastic diagonal view would take (1)–(3) as expressing their diagonal contents as well. But it is important to flag that it is possible to adopt a view on which these come apart.

> of the US in 2022 and supposes that Trump were in Beijing. She infers from those suppositions that person who is president of the US in 2022 in the actual world would be in Beijing.

Suparna's inference, by deductive standards, looks terrible. It isn't safe to infer from the counterfactual supposition that Trump is president in 2022 and in Beijing that the *actual* president from 2022 would thereby have to be in Beijing. Obviously the actual president in 2022 need not be, and in fact is not, Trump. Note again that this seems the appropriate verdict even once we are careful to separate out the question of whether reasoning is good from whether an inference made in the course of reasoning is good *qua* inference. Suparna's reasoning may be bad because she is wasting her cognitive resources on idle suppositions. But even if Suparna is not reasoning in the way she should, we can ask if the inference itself is performed correctly, just as we did for Belle. And the answer seems strikingly different from Belle's case. We should also note that the difference in the goodness of the inference can't trace to features of appreciability: we've stipulated both Belle and Suparna have the same capacity for appreciation. Finally note that, as in Belle's case, some logics can model Suparna's inference as a bad one. In particular, natural formalizations of Suparna's reported suppositions fail to be related by general consequence.

The following seem like plausible claims about Belle and Suparna's reported inferences.

(I) Belle's inference in (B) is a good deductive inference and Suparna's in (S) is bad.

(II) (B) and (S) report Belle and Suparna as bearing their attitudes to the same succession of contents.

But provided Belle and Suparna have, and exercise, the same capacities for appreciation in inference, these two claims are in tension with the following principle about deductive inference.[18]

> Uniformity.
>
> The goodness of a deductive inference always depends solely on a relation between the contents involved in the inference, and some cognitive grasp of that relation.

[18] For independent endorsement of this principle in a different context, see Valaris (2011).

These cannot all be true. By UNIFORMITY, the same standards of goodness govern Belle and Suparna's inferences. In particular, the goodness of those inferences depends only on the contents believed or supposed, up to facts about appreciability (or whatever plays the 'taking role'). But we are supposing Belle and Suparna have, and exercise, the same capacities for appreciability. So the goodness of Belle and Suparna's inferences depends only on the contents they believe and suppose. But by (II), the contents of Belle and Suparna's inferences are the same. It follows that either both of the inferences described in (B) and (S) are good, or both are bad. But this contradicts (I).

Note that the claims about the standards of goodness for inference in UNIFORMITY are neutral on the question of whether that goodness involves necessary truth preservation. All that is claimed is that the standard depends on the contents involved in an inference. The claim is even neutral on whether it is the truth-conditions of the content that matter—for example, structure may be relevant. In spite of this neutrality, UNIFORMITY continues to conflict with (I) and (II).

Denying UNIFORMITY comes with the obvious cost of complicating or fragmenting our conception of good inference. One option for denying UNIFORMITY is to continue to maintain that inferential standards are constant across attitude types, but depend on more than an appreciability requirement and the contents involved in the inference. This way of rejecting UNIFORMITY is under-motivated. Bracketing the issue of differing attitudes, it is hard to motivate an additional requirement on inference that goes beyond an appreciability requirement and a suitable relation between contents. When one person correctly infers q from $p_1, \ldots, p_n$, it seems that any other appreciated inference with those contents also counts as a good one. We could claim that there is a special extra condition that is only witnessed in cases like (B) and (S). But it is hard to see what that extra condition would be that didn't somehow trace to the difference in their attitude states.

So the more principled way to deny UNIFORMITY is to claim there are different standards of inferential goodness for belief and counterfactual supposition. This view is a more natural reaction to examples like (B) and (S). But it does face a continuing concern, which is that deductive inferences that appear to be good in the context of belief also appear to be good in the context of counterfactual supposition, and vice versa. Again, the divergence only seems to arise in special cases like those of (B) and (S). So this route will require positing

diverging standards of goodness for different attitudes that must nonetheless produce broadly similar results. Indeed, except in special cases, those diverging standards much produce exactly the same result. At the same time, the standards need to be different enough to distinguish (B) and (S). And, presumably, we need some kind of explanation for why the highly similar but ultimately differing standards for goodness apply to the different attitudes.

I do not think that these obstacles for rejecting Uniformity are insurmountable. But I do think the resulting account would feel gerrymandered in ways that motivate looking for an alternative. An account which maintains Uniformity is not only simpler but, as I'll argue now, has some important independent linguistic motivations. The account I have in mind embraces the disjunction I mentioned above: either Belle's inference is not good after all (so we should reject (I)) or the doxastic diagonal view is true (and can be used to reject (II)).

To motivate the disjunction, let's first see how embracing the doxastic diagonal view can help matters. Note that the doxastic diagonal view is only committed to diagonal content figuring in the belief complements in (B). So the diagonal view is compatible with saying the following: when we report an agent as making a counterfactual supposition that S, the report characterizes the supposition state of the agent involved using the horizontal content of S; and when we report that an agent believes that S, the report characterizes the belief state of the agent involved using the diagonal content of S. Perhaps it is the mood-marking in the suppositional complement which creates this differential effect, or perhaps it is accomplished by the semantics of the supposition report (suitably disambiguated from non-counterfactual supposition).[19] I'll remain neutral on this issue.

The proposed view gives the most natural and straightforward explanation of the badness of Suparna's inference. Assuming that we are working with horizontal contents, the truth of Suparna's basing attitude of counterfactual supposition is compatible with only metaphysically possible worlds in which Donald Trump is the president of the US in 2022 and is in Beijing. In some of those worlds Biden is in Beijing, and in some he isn't. The contents of Suparna's initial counterfactual suppositions don't settle that issue.[20] But the

[19] The issues here are complex. See Fintel & Iatridou (2023) for a recent discussion of the related morphological markings that distinguish 'counterfactual' from indicative conditionals, and their presence in related modal constructions.

[20] One might worry that the content of the counterfactual supposition is sensitive to facts

content of Suparna's concluding attitude of supposition does, or at least can, settle that issue. Since Biden is the actual president of the US in 2022, then supposing that the actual president were in Beijing would be to suppose something which would be true if Biden were in Beijing. Or, at least, since Biden is the relevant actual president, supposing that the actual president in 2022 is in Beijing *could* be to suppose Biden is in Beijing for all Suparna knows, since Suparna cannot know that Biden is not the actual relevant president (since she cannot know what is false). Either way, the restricted set of outcomes supposed to hold in the conclusion weren't ensured merely by the truth of the content of Suparna's basing attitudes. That, I submit, is what makes the inference problematic. If we want to help ourselves to this natural explanation, we must take Suparna's suppositions to relate her to the horizontal contents of the suppositional complements in (S).

Importantly, if this is the correct explanation of what went wrong with Suparna's inference, and we also assume as per (II) that the contents of Suparna's inference are the same as those involved in Belle's inference, then it becomes hard to see how Belle's inference could be any good. If Belle starts out by believing that the actual world is among some worlds which leave open some fact (say, whether Biden is in Beijing), and then deductively infers a conclusion which settles that fact, how can this count as a good inference? It seems like it can at best be a good ampliative inference—but probably not even that.

Things are different if we allow, as on the doxastic diagonal view, that (B) reports Belle as believing the diagonal contents of the relevant attitude verb complements. If that is the case, then the object of the belief that the 2022 president is in Beijing and the object of the belief that the actual 2022 president is in Beijing are true at exactly the same metaphysically possible worlds. On this hypothesis, (B) begins by reporting Belle as taking attitudes to contents whose truth leaves open the question of whether Biden is in Beijing. But Belle also concludes with a belief in a content whose truth continues to leave that question open as well, as the concluding content is true at many worlds where Biden

about the actual world, and if so that if Biden is not in Beijing in the actual world (say), it may be that in all worlds consistent with Suparna's supposition Biden also isn't in Beijing. Even if this were true, it doesn't get to the hear to the heart of the problem I want to raise. It seems Suparna should be able to suppose *some* additional content that would force the supposition-compatible worlds to leave Biden's location open. If this is done, the inference will continue to seem bad, while Belle's inference, augmented with the additional premises, will continue to be acceptable.

is not in Beijing. More specifically, the content is true at any worlds where the president of the US in 2022 is in Beijing, some of which are worlds where Biden is not president. (If desired one could add, building on ideas discussed in §8.1, that when we consider whether the contents of Belle's beliefs are true at a given, possibly counterfactual world, we are considering these worlds 'as actual' as we evaluate the content of belief. This would explain why the belief that the actual president of the US in 2022 is in Beijing is truth-conditionally equivalent to the belief that the president of the US in 2022 is in Beijing.)

Note: this view is emphatically *not* committed to the claim that the complements of the attitude verbs differing only in the presence of some form of 'actualization' behave the same way when they figure as the complements of (natural language) metaphysical necessity modals. This is not the case, and is precisely the bit of data that we saw Zalta try to exploit. But that interaction between the complements and necessity modals is separable from the question of what content is associated with the complements in attitude reports. Indeed, the view under consideration already maintains that the presence of actualizing language makes a large difference to complements when they figure as the object of counterfactual suppositions. Once we allow that the complements can sometimes express their horizontal contents, there is no obstacle to saying that this is what they do as the objects of natural language metaphysical necessity modals.

Note also that this account has safeguarded Uniformity. What makes Belle's inference good is precisely that it has (appreciably) preserved truth at all metaphysical possibilities. What makes Suparna's inference bad is that it has failed to meet that standard by failing to preserve truth at all metaphysical possibilities. There is just one standard governing inferences involving beliefs and counterfactual suppositions. It's just that natural language resources for expressing mental content are somewhat flexible, so that superficially similar sentences end up characterizing different attitudes with different forms of truth-conditional content.

This account retains a little bit of the idea, that would have been appealed to in denying Uniformity, that there is *something* about the attitudes involved in (B) and (S) that accounts for the difference in the goodness of the respective inferences. Belief in some loose sense 'aims' at actual truth. The worlds compatible with one's beliefs are treated as live possibilities for how things actually are. Counterfactual supposition does not behave in this way. Surely this

is a key part of the explanation for why, when we evaluate the truth-conditional content of a belief at a possible world, that world is 'considered as actual' in just the way that the doxastic diagonal view posits. And it is a key part of the explanation for why that assumption no longer holds when we evaluate the truth of the content of counterfactual suppositions. But we can maintain that the attitudes make a difference in (B) and (S) in this way, without going so far as to claim that the inferences they involve are governed by fundamentally different standards of goodness. Rather, since believers and counterfactual supposers have different cognitive aims, we use natural language complements in different ways to characterize the contents of their attitudes in accord with those aims.

So opting for the diagonal view affords us an intuitive explanation of why Belle's inference is good and Suparna's is bad, all consistently with the existence of a uniform standard for inferential goodness. The principal cost of this view is the denial of (II), required to maintain that the truth-conditional content of Suparna's suppositions and Belle's beliefs are different.

It is worth noting that there is some independent linguistic data which bears on this issue. There are other contexts in which the content of belief and ordinary supposition reports seems to diverge from that of counterfactual supposition reports, even though the reports involve similar attitude verb complements. The data concerns the behavior of negative existentials and, as far I as know, was first noted by Kripke in his 1973 Locke Lectures.[21] The data was notably exploited by Stalnaker in his classic defense of the utility of metalinguistic diagonal assertoric content.[22] Stalnaker's discussions centered on the difference between embeddings of negative existentials in the antecedents of indicative and counterfactual conditionals. But given the close ties between such conditionals and supposition states, the same data can be reworked to concern attitudes.

Consider the street artist who goes by the name "Banksy". Let us imagine we are speaking in a context in which it is an open question whether there really is a single person named "Banksy" (as opposed, say, to a collective of artists working under that name, or a group of diffuse and otherwise unrelated 'copycat' street artists producing the work that is mistakenly attributed to a single individual.) In that context, I instruct you to make a supposition as in (4).

[21] Kripke (2013).
[22] Stalnaker (1978, 329–32).

(4) Suppose that Banksy doesn't exist and that that name refers to someone else.

This request is highly perplexing. It has a kind of contradictory feel. If "Banksy" refers to some particular person, that seems sufficient for Banksy to exist. So how can one suppose Banksy doesn't exist while that very name succeeds in referring? These seems to be something similarly odd about a belief report as in (5).

(5) I think that Banksy doesn't exist and that that name refers to someone else.

The beliefs expressed in (5) seem incoherent or obviously false. But contrast (4) and (5) with an instruction to produce the corresponding counterfactual supposition as in (6).

(6) Suppose Banksy hadn't existed and that that name referred to someone else.

This seems readily supposable. For example, some have hypothesized that Banksy is Robert Del Naja, the frontman for the trip-hop group Massive Attack. In this case, one way to fulfill the instruction in (6) would be to suppose that Robert Del Naja was never born, but that someone else took up Banksy's now iconic moniker (perhaps to produce artwork of a similar kind).

Here is what is especially interesting about this contrast. I've just described a possible world compatible with what one would suppose in supposing as per (6). But supposing the world is that very way does *not* seem like a way of supposing that would comply with (4). For example, to suppose that Robert Del Naja doesn't exist, and the name "Banksy" belongs to someone else does not seem like a way of complying with (4). It is just to suppose that Del Naja isn't Banksy after all. Nor does thinking the world is this way appear to be a way of thinking that would be properly reported with (5).

If these are more than mere appearances, the truth-conditional content of the complements in (4)/(5) and (6) must diverge: there are worlds at which the content expressed by the complement of (6) are true but at which the content expressed by the complements of (4) and (5) are not. Why do the contents diverge in this way? Well, the content attributed in (6) appears to be the customary horizontal content of the verb complement. This is the content which

holds fixed the actual referent of "Banksy" and evaluates a world for truth according to whether that individual fails to exist while someone else bears his name. That is why supposing the possibility I discussed for Del Naja (provided he really is Banksy) would count as compatible with that content. But this would mean that the content reported as believed in (4) and supposed in (5) was not this horizontal content. We can instead take this to be a particular form of diagonal content known as the sentence's *metalingustic* diagonal content. This, very roughly, is the content which would be true at a world just in case the sentence were true as used at that world.[23] The sentence "Banksy doesn't exist and that name refers to someone else" is never true as used at any particular world for precisely the reasons given above: if it is true to assert that the name "Banksy" refers, then it is not true to say "Banksy doesn't exist" using that very name.

The evidence from (4)–(6) seems important to consider when evaluating reports like (B) and (S). It shows that we have reasonably strong intuitions from a range of contexts about the divergence in truth-conditional contents between belief reports and supposition reports about actuality on the one hand, and counterfactual supposition reports on the other. What is more, those intuitions tend to track something like split diagonal/horizontal contents in precisely the way that would allow us to safeguard Uniformity.

But here, as noted above, I will stop short of endorsing such a position. This is because the data from (4)–(6) (and accordingly (B) and (S)) is inconclusive. There is important further data that complicates the picture. For example, intuitions about asymmetries between belief and counterfactual supposition seem to disappear when we consider third-person reports.[24]

(7) Mark believes that Banksy doesn't exist and that that name refers to someone else.

It is possible to hear (7) as true in cases when Mark knows Robert Del Naja, but not under the name "Banksy", believes that Del Naja is merely a figment

[23] See Stalnaker (1978) for more details. Importantly, on Stalnaker's view we hold some semantic features of the sentence—those we presuppose the words to have—fixed as such truth-conditions are determined. If there are words whose semantic properties are left open by what we presuppose, then we instead allow the semantic properties of those words to *vary* within the range of properties we presuppose them to have. In this case: we know the semantics of all terms in the complement, but do not know the referent of "Banksy", which is accordingly allowed to vary as we assess different worlds. This is what produces the relevant truth-conditions.

[24] A fact again noted by Kripke (2013).

of his imagination, and thinks the name "Banksy" belongs to someone else. If that is right, then there is pressure to treat the complement here as expressing the horizontal content of the complement again.

How to respond to the total data set in (4)–(7) is a very complicated matter. For example, one could maintain a further split among belief reports, on which some take as objects their complement's diagonal content while others take as objects their complement's horizontal content. Perhaps context would do the work of disambiguating, either covertly in the complementizer phrase, or in interaction with the attitude verb itself.

But one could also maintain that (7) shows that our intuitions about (4) and (5) should be explained away by an error theory. For example, one could maintain that (5) can be a true report made by a thinker who coherently believes what is expressed by the verbal complement in (5), but that any thinker making such a report would nonetheless be irrational. Here is one way this explanation could begin: perhaps in believing Robert Del Naja doesn't exist, and that the name "Banksy" belongs to someone else, one *is* believing coherently in a way that would make the report in (5) true. But it may be that one couldn't rationally be in a position to report that belief by using (5). To do so, one would need reason to think "Banksy" referred to Robert Del Naja in order to recognize that one's beliefs about Del Naja would count as satisfying the relevant content expressed by the verbal complement. But recognizing *that* would precisely contradict what is reported in the second conjunct. This would assimilate (5) to standard explanations of why the Moorean "*p* and I don't believe *p*" is unassertable. Even though such statements are unassertable, the content asserted can clearly be true.

There is another way of maintaining the error theory. One could take it to be the case that a sentence like (4) or (5) can be used to simultaneously assert two distinct propositions. One proposition—the literal content of the sentences—is the one that corresponds to giving the complements their horizontal content. But another proposition asserted—one that is not the literal content, but is otherwise assertorically conveyed—is something closer to that associated with the diagonal view.[25]

I don't want to pursue these various lines of thinking any further. What I want to note for now is that while the data given so far doesn't obviously

[25] See Soames (2005) for a view with elements that start to take this shape, though Soames does not, to my knowledge, consider these examples specifically.

suggest one particular way of dealing with (4)–(7), it does provide reasonably strong evidence for a *disjunction*. We have some important evidence that our first blush intuitions about reports like (4) and (B) are responsive to something like the diagonal conception of content. What remains contested, in light of what I've discussed so far, is whether or not that diagonal content *actually* figures as the content reported as an attitudinal object in the literal assertions of the relevant sentences. But for my purposes, this open question is not of highest importance. However we answer it, we will have resources to apply to the case of (S) and (B). Recall the two theses which were applying pressure to Uniformity.

(I) Belle's inference in (B) is a good deductive inference and Suparna's in (S) is bad.

(II) (B) and (S) report Belle and Suparna as bearing their attitudes to the same succession of contents.

If there are genuine diagonal/horizontal asymmetries between (4)–(5) and (6), we have good reason to think such asymmetries also hold of (B) and (S). If so, we are in a position to deny (II). If, by contrast, the attitudinal diagonal view fails, it seems that the best explanation for why it fails appeals to an error theory, on which our intuitions about the truth or assertability of certain sentences like (4)–(5) hinges on mistakenly attributing diagonal contents to sentences or verbal complements which have horizontal content. We confuse what the sentences literally assert with what they convey by other means, or what it is reasonable to infer from their truth, and so on. If this holds of (4)–(5), however, there are good reasons to think the same problem afflicts (B). And if this is the case, we then have the resources to deny (I): our intuitions about Belle's good inference will be derived from our first blush misconstrual that the sentences report attitudes taken to diagonal content. There would be a good inference in the offing—it just wouldn't necessarily have to be the one literally expressed by the sentences in (B). And that good inference could still count as good precisely because it is necessarily truth-preserving, since that is the key property of the diagonal content transition that is (perhaps misleadingly) informing our intuitions.

What this shows is that as long as we can defend either the claim that the attitudinal diagonal view is true, or the claim that our intuitions are based on misattributed attitudinal diagonal content, we have the resources to reject (I)

or (II) and thereby safeguard UNIFORMITY. So far I've tried to explain why I think there is evidence that speaks in favor of that disjunction. Even so, I recognize that I am far from having settled this issue in this short series of remarks. The data I've looked at must be examined more carefully, and even more data compiled. And once the data is examined we must consider the full range of options for responding to it. In particular, we will have to weigh the disjunctive view I've defended against the virtues of any view that denies UNIFORMITY by splitting up the standards for inferential goodness belonging to belief and counterfactual supposition. I think that there are versions of that maneuver that are well worth considering, though I won't try to elaborate any particular attempt here.

Rather than pursuing the matter further, I want to step back and reflect on some broader lessons of the discussion so far. First, I want to make a remark about how a mix of theoretical and empirical questions in linguistics have come to bear on the shape—indeed the very foundations—of logic. Second, I want to explain why I think that although these questions introduce a great deal of complexity into logical debate, it is that very complexity which shows that we can make interesting forms of progress in settling logical matters.

In this section, I began with a series of reports in (B) that raised concerns about my account of inference. I defended against those concerns by accumulating more data, and trying to motivate a disjunctive view that appealed to controversial commitments about the compositional semantics of attitude reports and the nature of assertoric content in English.

Now, in one way it should come as no surprise that empirical matters could enter into our choice of logical framework. After all, I've claimed that logics modeled on a natural language will have to be responsive to the correct semantics for that language. And uncovering the semantics for a natural language is, in large part, an empirical enterprise.

But the role of empirical questions about language in this section goes much deeper than this. If the only question raised by (B) was about the correct semantics for English, we could have safely stopped our discussion at the end of §8.1 and let the empirical chips fall where they may. But this was not the only question raised by (B). (B) posed a challenge to the very conception of inference that I used in Part I to characterize the nature and purpose of logical inquiry. I developed that conception of inference, like so many philosophers before me, on the basis of basic judgments about when agents counted

as performing inferences, which kinds of content figured in those inferences, and when the inferences counted as good. But judgments about the mental contents involved in inference are heavily informed by the language we use to convey or report them. And our understanding of our own language can be imperfect. Moreover, as we've just seen, a mix of empirical and theoretical considerations may complicate our judgments of whether or not a good inference has been performed. By doing this, those empirical and theoretical considerations may complicate our understanding of inference itself.

This shows my conception of logic to be embroiled in controversy. If what I've been arguing in this section is true, a 'correct' logic may depend on a range of vexing issues in the philosophies of language and mind. Settling how to cope with sentences like (S), (B), and (1)–(7) may depend on the tenability and proper interpretation of two-dimensional semantic frameworks, on the distinction between what is literally said or asserted and what is pragmatically conveyed, and possibly on the proper response to Frege-style puzzles in which one object is known under two 'guises' (since it seems like such cases are the ones on which we most naturally get true readings of (7)).

But, in keeping with the motivation for this book, I think some amount of controversy like this should be welcome in the foundations of logic. The question of whether logical truths are metaphysically necessary is an important one. It would be disappointing is if the controversy surrounding that question bottomed out in brute intuitions about correct inferences or logical frameworks. It would be equally disappointing if it led to the kinds of simple dead ends that we found when considering the positions of Zalta and Hanson. What we can see now is that even this high-level debate in the foundations of logic about whether logical truths are necessary can be tethered to concrete problems in non-logical domains like empirical semantics. These problems are responsive to much more than intuitions about 'what (logically) follows from what.' And even though the issue is hardly settled, we have a good sense about how to proceed in taking the next steps to resolve it.

As acknowledged, I haven't said enough to decisively defend my approach to the puzzle about inference raised on the basis of Zalta's case. Still, enough has been said, I think, to justify a continued methodological focus on the logical study of good deductive inference construed as requiring the metaphysically necessary preservation of truth among inferential contents. Apparent obstacles to treating inference in this way from sentences like (B) are much more

complicated than they initially appear. There are empirical considerations that help to explain away apparent problems, as well as technical resources to codify those explanations. We should bear in mind, of course, that this is mere methodological advice. The tenability of my approach is unavoidably tied to tricky theoretical and empirical questions that cannot be resolved here. But I wouldn't have things any other way. Controversy in logic is unavoidable. What we want is *direction* in how to resolve such controversy that doesn't bottom out in question-begging brute intuitions. One of my aims in discussing the problem cases of this chapter is to show one way in which my approach to logic can give us some direction of this kind.

CHAPTER 9

Validity in the Presence of Semantic Defect

In Chapter 7, I argued that to produce a genuine violation of Excluded Middle we would minimally need to make foundational sense of a third truth-value in application to possible assertoric contents of sentences (or sentences themselves, *qua* vehicles of assertoric content). We might also need empirical arguments that showed this value arises within some fragment of discourse of interest to us, and that it projects compositionally in certain 'infectious' ways (that is, tends to be inherited by truth-evaluable compounds from their truth-evaluable parts). Having flagged the importance of these issues belonging to the philosophy of language and linguistics respectively, I left them unsettled, as my goal was primarily to show how the question of whether Excluded Middle holds turns on substantive questions that do not prejudge the truth of that putative logical law.

Still, even if we can't resolve these kinds of issues here, it is well worth considering an important hypothetical: what *would* a logic look like if an infectious third truth-value pervaded some sphere of discourse we wanted to investigate in logical terms? Answering this question is the aim of this chapter.

I begin fixing ideas in §9.1 by discussing two common alleged sources of semantic defect in natural languages along with their customary motivations. I then discuss two lessons we would learn from developing a logic for either kind of defect. The first is that although defect tends to 'weaken' a logic, in the sense of permitting fewer logical entailments, this apparent weakening is better understood as a process of rebranding some familiar instances of logical entailment as instances of *general* entailment. In particular, logics of defect do not treat formerly recognized good inferences as bad inferences, but instead as good inferences of a non-logical kind. The second lesson concerns the (often implicit) ways that a consequence relation is relativized to a stipulated set of lin-

guistic properties. Semantic defect forces us to get clearer on what is involved in this kind of relativization, and reveals a salient choice about which kinds of linguistic properties we should focus on in a logic for defect. This produces two quite different alternatives for developing a consequence relation in the presence of semantic defect one of which, perhaps surprisingly, is none other than classical logic.

In §9.2, I go on to apply these lessons to a form of objection, commonly made in the context of developing formal theories of truth in response to the liar paradox, that logics governed by the strong and weak Kleene schemes are too weak to carry out 'ordinary reasoning.' I note that both lessons of §9.1 problematize such claims, especially when they are directed against theorists who are relatively clear about the nature of the semantic defect perturbing logical relations in their systems. Saul Kripke, in his contributions to formal theories of truth, will provide an example of such a theorist. While there are avenues to pursue to try strengthening the original objections by appeal to the peculiarities of 'contingent semantic defect' (connected with the phenomenon of 'contingent paradox'), even the success of such further appeals are of dubious utility once it is recognized how many resources the theorist of semantic defect can appeal to in accounting for ordinary good reasoning in such cases. Without space to pursue the dialectic further, I rest content with the conclusion that, as they currently stand, objections to weak logics in the context of theories of truth are underdeveloped and so (as of yet at least) unpersuasive.

I conclude by extracting a general lesson in §9.3, which is that it is generally a mistake to object to the weakness of a deductive inferential logic on the grounds that this weakness threatens our capacity to reason. This is because such weakness can seemingly at most reflect problems for the logician as theorist arising from the complexity of their target subject matter, not problems for any reasoners whose inferences a logician is in the business of modeling.

9.1 Infectious, Inference-Blocking Defect

How would the presence of semantic defect influence logic? Getting clearer on the answer to this question requires addressing some further issues: What is semantic defect? When does it get inherited through embeddings? And most importantly, what bearing could semantic defect have on the study of good inference?

Let me begin by drawing a customary distinction between two kinds of semantic defect that can belong to declarative sentences. First, such a sentence may fail to express a proposition (that is, fail to express a mind- and language-independent truth-evaluable object of attitudes). Second, it may succeed in expressing a truth-evaluable proposition that nevertheless fails to be true or false relative to at least some possible worlds.

The foundational coherence of the second of these two notions is subject to contention. Beginning with Dummett, philosophers have wondered what sense there could be in applying a third status beyond truth and falsity to assertoric content.[1] By contrast, the coherence of the claim that sentences could fail to express propositions is not in serious dispute. After all, plenty of other things (e.g., antlers or yams) clearly cannot be used in normal contexts to express propositions. The only interesting question remaining is accordingly the empirical one of whether any grammatical declarative sentences of actual natural languages do fail to express propositions.

Why posit either of these kinds of defect? A starting point is usually truth-value intuitions about sentences. Definite descriptions with unsatisfied descriptive material, as found in (1), seem to exhibit a form of infelicity that many speakers are inclined to describe as leading to the sentence's being neither true nor false.

(1) # The present king of the United States is in Germany.

This is connected with the fact that the infelicity tends to be preserved under negation. (2) seems to be problematic in the same way as (1).

(2) # The present king of the United States is not in Germany.

Both of these ideas are of course mere points of departure. Speaker intuitions of truth-evaluability are notoriously unstable.[2] And as we've already had occasion to discuss in Chapter 8, it is important to mark a conceptual separation between a sentence's assertoric content and various compositional questions, such as how sentences behave in embeddings.

It is worth contrasting the defects of non-referring descriptions with non-referring complex demonstratives.[3]

[1] For other noteworthy skeptics see Williamson (1994) and Glanzberg (2003).

[2] See, e.g., von Fintel (2004).

[3] See Glanzberg & Siegel (2006) for an extended discussion of complex demonstratives with unsatisfied nominals.

(3) # That terrier [pointing at a siamese cat] is dangerous.

Although (3), just like (1), may result in intuitions of truth-valuelessness as well as inheritance of sensed defect under certain embeddings, the fact that it involves a complex demonstrative rather than a description gives grounds for a significantly different semantic treatment. Descriptions like "the present king of the United States" are typically treated as non-rigid, in contrast with demonstratives and complex demonstratives. As such, even if "the present king of the United States" fails to refer at the actual world, it may refer to an individual at counterfactual possible worlds in which the United States became a monarchy. That seems to imply (1) can express something—a proposition—that is true or a false in other worlds. By contrast, if "that terrier" in (3) lacks a referent, it arguably lacks that referent at all metaphysical possibilities.

The semantic differences between (1) and (3) may go deeper than just the modal profiles of their assertoric contents. Demonstratives are typically treated as 'directly referring' in something like the sense of Kaplan (1989b): the contribution to assertoric content made by the expression is exhausted by its referent. If complex demonstratives also behave this way, and fail to refer when they have unsatisfied nominals, then there will simply be no contribution to assertoric content made by the complex demonstrative in (3). And that may lead to the view that (3) expresses no proposition—no possible object of assertion or belief—at all.[4]

This contrast leads to differences in embeddings in indirect speech reports. Contrast an assertion of (3) with one of (4) by the same speaker, and associated 'echoic' indirect speech reports (5) and (6).

(3) # That terrier [pointing at a siamese cat] is dangerous.

(4) # The terrier I'm now pointing at [pointing at a siamese cat] is dangerous.

(5) # Pia said that that terrier [pointing at a siamese cat] is dangerous.

(6) Pia said that the terrier she was pointing at is dangerous.

Any infelicity exhibited in (4) needn't be preserved in at least one reading (the de dicto reading) of (6). If (6) can be true, it seems like there is something

[4] Though for an alternative possibility, consider the view that these involve 'gappy' structured propositions of the sort advocated by Braun (2005).

Pia said with (4)—that is, some proposition she expressed. By contrast, any infelicity in (3) seems to be inherited by a report like (5). This may increase the sense that there isn't anything that Pia said with (3)—that is, that she expressed no proposition with those words.

The foregoing discussion has merely been aimed at fixing ideas: to note the conceptual differences between two potential types of semantic defect and the kinds of data that one might marshal in favor of treating some sentences with one type of defect or the other. Let's turn to the question of what implications the presence of these forms of defect would have for logic.

The first thing to note is that either form of defect would interfere with good inference. Take first the case of propositional expression failure. As I argued in Chapter 2, good deductive inference aims at a correctness-preserving transitions of acceptance states. If a sentence fails to express a proposition, then there is nothing for a concluding acceptance state to take as an object. Accordingly such a sentence can't be used to model an inference at all, let alone to model a good one.

This outcome is not inevitable. It may be avoided by views that allow for mental analogs sentences, considered as mere vehicles of content. For example, on the Language of Thought Hypothesis, contentful thought takes place in a kind of mental symbolism with a structure that mirrors that of natural languages.[5] On such a view, someone asserting a sentence like (3) may have a mental analog of that sentence physically realized in their brains. It is open to take that physical realization as a *kind* of concluding attitude state of acceptance, albeit one which lacks content.

But even if there is a concluding attitude of this kind, it should be undisputed that any such state cannot be regarded as correct (relative to a world) in the sense that bears on good inference. Correctness of this kind marks something like the compatibility of the information contained in the information state with a way the world is. And on the current hypothesis there is no such information to begin with.

So whether sentences that fail to express propositions correspond to attitude states which can mark the endpoint of an inference, it should be clear that they cannot be used to represent the endpoints of *good* inference in logical inquiry. Any such endpoints are governed by a standard of correctness assessed

[5] See Fodor (1975) for a seminal exposition of the idea, and Rescorla (2019) for an overview of the relevant literature.

on the basis of informational content, and there is no such content expressed by the sentence to begin with.

What about sentences that express trivalent content? There is room for a position which treats the worlds at which trivalent propositions exhibit defect to behave as 'local' failures in the expression of information-bearing properties of attitudes. On this view, even though a trivalent proposition can be the object of attitudes, the third truth-value marks worlds relative to which the trivalent proposition can do no work in characterizing the structure of a mental state.[6] If this view held, defects in trivalent propositions would sometimes block good inferences for the same reasons as propositional expression failures. It's just that they would do this only relative to the worlds where the defects arise. But we needn't take such a controversial stance to give trivalent propositions good-inference-blocking statuses. As before, all we need to presume is that the information contained in an attitude which takes trivalent content as an object are not compatible with the worlds where truth-valuelessness arises. There may be alternatives which do not treat a third value with that status, but it is hard to see why such views would be counted as forms of semantic defect to begin with.

So to sum up, trivalent defect and propositional expression failure both interfere with inferential goodness. They may do this for one of many different reasons: because they mark the absence of possible attitudes, or the presence of attitudes without content, or the presence of attitudes with 'partially characterizing' content, or the presence of attitudes with informational content merely incompatible with worlds at which that content is defective. Though these sources of inferential interference are different in important respects, it is not clear that logic is in the business of minding those differences. Accordingly, we can simply lump all these forms of defect together from here on out as good-inference-blocking.

In this way, semantic defect so far has no more special bearing on logic than does falsehood. Both are good-inference-blocking statuses. What is typically thought to be logically interesting about semantic defect is its 'infectious' character: its tendency to be preserved under embeddings in ways that falsehood is not. This idea arose above in discussing how the infelicities of definite descriptions with unsatisfied descriptive material, and complex demonstratives with unsatisfied nominals, tend to be preserved under negation. Falsity is not

[6] See Shaw (2014).

preserved in this way.

I will proceed on the assumption that semantic defect is infectious in the aforementioned sense. But it is worth nothing that this is an assumption that would eventually need argument. All the forms of defect I've described so far are defects that arise at the level of assertoric content (even if the defect is failure to express such content). This has to be so if the defect is to have any bearing on logic. For only defects at the level of content have any bearing on the states that inference mediates between. But as we've seen in Chapter 8, it is important to keep claims about such content distinct from claims about compositional semantic values. It *may* be that certain kinds of defect tend to lead to certain compositional effects. For example, it may be that if a sentence fails to express a proposition, then sentences embedding it must likewise tend to fail to express a proposition. But however natural such claims may seem, they require argument. And this justification is especially pressing for the other kinds of defect that I've described, which don't involve propositional expression failure.

Though argument would eventually be needed to connect semantic defect and inheritance behavior, I will proceed for now on the assumption that the connection exists. The reason for this is simple: if we can argue for semantic defect that has the inheritance profile of something like falsehood, then it is clear that almost no interesting consequences arise for logical inquiry. By contrast, an infectious form of good-inference-blocking defect could have striking ramifications.

To see why, let's suppose that some form of defect can arise in a first-order language (or a language that is regimented in a first-order setting) at the level of predication. We can model this possibility in a familiar way, by associating predicates with extension/anti-extension pairs, so that predications of objects in the extension yield truth, predications of objects in the anti-extension yield falsehood, and predications of objects in neither set yield a 'gap' value, representing some form of defect. Let's further suppose this gap value has the most infectious behavior possible—that represented by the weak Kleene scheme. (Though I focus on this scheme for simplicity, the lessons we extract will generalize naturally to other schemes such a the strong Kleene scheme, etc.) On this scheme, any truth-functional compound (a conjunction, disjunction, conditional, or negation) possesses a gap value if any of its immediate constituents do, and any quantified expression possesses a gap value if at least one of its instances do.

Logical consequence relations on the view I'm exploring track relations of necessary truth-preservation that hold in virtue of some type of linguistic properties. In Chapter 7, I defined a a type of linguistic property—modalized first-order form—such that first-order consequence relations capture necessary truth-preservation in virtue of the possession of those properties. But modalized first-order form included the property that predication was necessarily bivalent, and that the values of connectives and quantifiers were determined from bivalent values in the ordinary way. Many sentences of our newly hypothesized language thus obviously cannot (even on modalized interpretations) have modalized first-order form in that sense.

The natural next step is to explore a consequence relation relativized to a new set of linguistic properties that relaxes the requirement of necessary bivalent predication. We can accomplish this by replicating the steps we took in Chapter 7, but transposed to the weak Kleene setting. A *gap model* for a first-order language is defined as a familiar classical first-order model, except that in it predicates are assigned an anti-extension in addition to an extension (we ignore functions for simplicity). An extension tracks objects such that predication of them would yield truth; an anti-extension objects such that predication yields falsehood; all other objects in a domain are such that predication yields a gap-value. We inductively define satisfaction relations accordingly, using the weak Kleene projection scheme to track inheritance, and then use the satisfaction relation to define what it is for a sentence to be true, false, and gap-valued in a gap model in the obvious way. We define a *modalized gap interpretation* for a first order language to be a function from metaphysically possible worlds to gap models, whose domains are drawn from the world to which the model is assigned. Using this final notion we can define a new batch of linguistic properties, giving a form common in weak Kleene interpretation.

> The *weakened modalized first-order form* (wMFOF) of a first-order sentence ϕ consists in the set of syntactic and base semantic properties shared by ϕ on all of its modalized gap interpretations given weak Kleene projection.

> The *weakened modalized first-order form* (wMFOF) of a set of first-order sentences S consists in the set of syntactic and base semantic properties that sentences of S share on all of their modal-

ized gap interpretations given weak Kleene projection.[7]

These definitions essentially take ordinary modalized first-order form and replace any base semantic properties associated with necessary bivalent predication with necessary trivalent predication, while also adding base semantic properties for compositional processes that correspond to weak Kleene inheritance for gap values.

With a new batch of linguistic properties specified in this manner, we are able to formulate new 'true' versions of logical validity and consequence.

> A first-order sentence type ϕ is *wMFOF-valid* iff necessarily, on any interpretation of ϕ that gives ϕ weakened modalized first-order form, ϕ expresses a necessary truth.

> A first-order sentence type ϕ is a *wMFOF-consequence* of a set of first-order sentence types Γ iff necessarily, on any interpretation that gives $\Gamma \cup \{\phi\}$ weakened modalized first-order form, every world at which all of the sentences of Γ are true is a world at which ϕ is true.

What would a logic investigating necessary truth or necessary truth-preservation in virtue of weakened modalized first order form look like?

Well, it would be a weak logic, in a familiar sense of licensing many fewer 'logical' inferences than the classical setting. To see the familiar idea, let's consider disjunction introduction. Consider first-order sentences (7) and (8) which, let us suppose, receive intended modalized gap interpretations that share a modal profile of truth conditions with (7′) and (8′) respectively.

(7) Ws

(7′) Snow is white.

(8) $Ws \vee Gg$

(8′) Snow is white or grass is green.

[7] Note: we need this separate definition in terms of sets of sentences to capture a notion of form relevant to a consequence relation, which I would like to explore here. In particular, we must look to base properties which *link* sentences (e.g. the property that a predicate letter receives the same extension across *two* sentences) to get the right characterization of entailment holding in virtue of form.

The transition from (7) to (8) is not one secured by weakened modalized first-order form. The predicate G and constant symbol g could be interpreted in a gap model at some world so that the predication Gg produces a gap while Ws continues to be true. Given the weak Kleene scheme, (7) would be true on that gap model while (8) would have a gap value. The existence of a modalized gap interpretation using this gap model at a world suffices to show that the transition from (7) to (8) does not necessarily preserve truth in virtue of weakened modalized first-order form.

This much should be straightforward. But two points are in order. The first is that claiming that (8) does not logically follow from (7) in this context *does not mean that an inference from (7) to (8) cannot be a good deductive inference*. On the contrary, supposing that (7) and (8) are *bivalent at all worlds on their intended modalized interpretations* (or even just bivalent at all the same worlds) would result in (7) semantically entailing (8). The occurrence of entailments that are non-logical is not new in this context. For example, lexical entailments (e.g., from "A is a vixen" to "A is a fox") fail to be logical entailments for essentially the same reason: although (typical) 'logical' linguistic properties are insufficient to guarantee a lexical entailment, further 'non-logical' semantic properties borne by sentences can pick up that slack. What is new in a language with highly infectious defect is mainly the degree to which such non-logical entailments can proliferate. If there are many necessarily bivalent[8] sentences of an interpreted language which *could* have had gap-like behavior on rival interpretations, that may suffice to rule out their logical truth. But it may have no *further* bearing on whether inferences without premises to the contents expressed by the sentences are any good. In effect, the property of 'being necessarily bivalent' is no longer among our 'logical' properties, *and as a result becomes a frequently possessed 'non-logical' property* to sometimes restore the goodness of an entailment.

The second point concerns the weakness of the logic. It is true that the logic of inference-blocking, infectious semantic defect can be weak in the sense of allowing substantially fewer logical entailments, as wMFOF-consequence shows. There are many fewer logical wMFOF-validities than MFOF-validities, and many fewer sentences related by wMFOF-consequence than by MFOF-

[8] Necessity here is quantifying over worlds, not interpretations. So I mean: "is true or false at all worlds on its intended modalized gap interpretation." Not: "is true or false on any gap interpretation."

consequence. For example, the wMFOF-consequence relation will tend to fail in any instance where a consequent introduces 'new material': applies a new predicate, or introduces a new term. Both such additions occur in our transition from (7) to (8).

But it is important to acknowledge that this weakness depends on how we characterize and apply the consequence relation, and in particular on the set of linguistic properties that we relativize the consequence relation to. So far, we've examined relativizing validity and consequence to weakened modalized first-order form. But this is not the only way to define a consequence relation in the trivalent setting. We *could* consider definitions of validity and consequence that potentially apply to first-order sentences bearing noteworthy semantic defect, but restrict the attention of the consequence relation to those sentences which bear the property of necessary truth-evaluability. Indeed, some investigations of logics in the trivalent setting have done precisely that. They are formulated in a manner analogous to the definitions below. As a nod to the existing usage, I'll term the resulting characterizations forms of 'Strawson' validity and consequence.[9]

> A first-order sentence type ϕ is *Strawson wMFOF-valid* iff necessarily, on any interpretation of ϕ that gives ϕ weakened modalized first-order form, *and* renders ϕ truth-evaluable at all worlds: ϕ expresses a necessary truth.
>
> A first-order sentence type ϕ is a *Strawson wMFOF-consequence* of a set of first-order sentence types Γ iff necessarily, on any interpretation that gives $\Gamma \cup \{\phi\}$ weakened modalized first-order form *and* renders them truth-evaluable at all worlds: every world at which all of the sentences of Γ are true is a world at which ϕ is true.

(8) is a Strawson wMFOF-consequence of (7). Though there are reinterpretations of the predicate G and that give the predication Gg a gap value relative to some worlds, those interpretations are by stipulation irrelevant to whether Strawson wMFOF-consequence holds.

[9] The terminology owes to von Fintel (1999), who is hearkening back to Strawson (1952). Previous forms of 'Strawson entailment' tend to concern individual extension assignments, and so not truth-value assignments at multiple possible worlds. Otherwise the connections between 'traditional' definitions of Strawson entailment and my versions here should be apparent.

The strategy used in the Strawsonian definitions of course comes with a cost. The cost is that some Strawson wMFOF-validities can be interpreted in ways that would not render them true (let alone necessarily true), and some sentences are related by Strawson wMFOF-consequence can be interpreted in ways that would not preserve actual truth (let alone truth at all worlds). For example, all instances of Excluded Middle are Strawson wMFOF-valid. But it is easily possible to interpret (or reinterpret) those sentences so that strong, infectious defect gives them a gap value. In fact, if we allow the definition to apply to interpreted as well as uninterpreted types (as I am allowing), there will be Strawson wMFOF-validities that are *actually* (i.e. on their received interpretations) false.[10]

In this way, the Strawsonian definitions are tracking a *syntactic* form that is *conducive* to necessary truth, or truth-preservation, without being sufficient for it. We actually saw an analogous instance of this kind of phenomenon in Chapter 7 in the discussion of the existence presuppositions built into modalized first-order form. There I noted that if (perhaps *per impossibile*) it were possible for nothing to exist, some MFOF-validities would not express necessary truths. MFOF-validity tracks sentences that *would* express a necessary truth in virtue of possessing a range of semantic properties, *were* they to possess them. This range of properties included the semantic property that any quantifiers necessarily quantify over a non-empty domain. If it were possible for nothing to exist, then some sentences like $(\exists x)(x = x)$ would vacuously count as MFOF-validities, since attributing modalized first-order form to them would require them to bear property that, necessarily, no quantified expression could. As a result, there could be ordinary, first-order validities in this sense that one could not safely infer from no premises. Indeed, there could be validities that

[10] This could be avoided by redefining validity and consequence in non-conditional terms, and only to *interpreted* sentence types. That is, rather than saying an interpreted *or* uninterpreted sentence type ϕ is L-valid if it would express a necessary truth were it to have had the properties of type L, we could say that an interpreted sentence type ϕ is L-valid just in case any sentence with ϕ's syntax would have expressed a necessary truth were it to had have the properties of type L, *and* ϕ *in fact* has those properties on its given interpretation. This redefinition has its own costs of course. Now the 'form' of a sentence ϕ cannot be read off of its syntax even given the relevant linguistic context. For example, some syntactic instances of Excluded Middle will be Strawson valid, and others will not, and the only difference will trace to semantic features the sentences could have possessed or lost based on reinterpretation. I have nothing against these kinds of definitions (see below for the ecumenical stance on characterizing validity). The only important thing is to be clear about how the logical classification is to be applied, and what its consequences are.

were false.

The difference between the example of $(\exists x)(x = x)$ for MFOF-validity, and the example of Excluded Middle for Strawson wMFOF-validity, is that attributing Strawsonian weakened modalized first-order form to an instance of Excluded Middle attributes a property to it that it merely *contingently* could lack. Otherwise, Strawson wMFOF-validity also tracks expressions that *would* express necessary truths *were* they to have a particular form. So, in this case, we can find interpreted instances of excluded middle that have a gap value, but nonetheless count as a Strawson wMFOF-validities because the sentences *would* have expressed a necessary truth *had* they possessed a range of properties including that of being necessarily bivalent.

To understand the significance of this feature of the Strawsonian definitions, we should turn to a different question: how many classical validities and consequences (i.e., MFOF-validities and -consequences) that were discarded in the transition to wMFOF-validity and -consequence have now be reintroduced in the Strawsonian variants? The answer is (perhaps obviously): all of them. In fact, as may well have been clear to the careful reader, Strawson wMFOF-validity and consequence *just are* MFOF-validity and consequence. By this, I don't merely mean that these notions are coextensive. Rather, they are *essentially the same* property and relation, defined in almost exactly the same way.[11] The modifier "w" ("weakened") in wMFOF signaled that, unlike in MFOF, we were accommodating the possibility of violations of necessary bivalence. The "Strawsonian" modifier simply undid this modification, by reintroducing necessary bivalence as a property to which our validity and consequence relations were relativized. The result is that we are back with the linguistic properties (and so validity and consequence relations) with which we started.

Why go through the trouble of subtracting and adding the selfsame properties to our characterizations of true forms of validity? I think that doing so can help hammer home the importance of a lesson that we encountered in Chapter 7: that the simplicity and well-behaved character of classical logic is in large measure a *stipulated* feature. What I want to bring out here is that the simplicity of the classical setting (at least insofar as this is analyzed as MFOF-

[11] 'Essentially' and 'almost exactly' since technically the above definition gets us to something like the properties of first-order modalized form in two steps: by attributing properties securing trivalence and then tacking on properties securing bivalence. In the definition of first-order modalized form we get there in one step: by attributing bivalence immediately. This difference is clearly inconsequential.

validity) is stipulated *in the very same sense* in which the Strawsonian simplicity is stipulated. To say a first sentence is a classical consequence of a second, and to say that the first sentence is a Strawsonian consequence of the second is, again, to say exactly the same thing.

I stress these ideas because I think there is a strong temptation to think that the Strawsonian definitions of validity and consequence are somehow gerrymandered or evasive. "Of course," one might say "Strawsonian validity and consequence appear well-behaved. But this is only because those definitions effectively operate under a pretense that every sentence is (necessarily) truth-evaluable, when they are not. Things may look simple when one misleadingly stipulates away the complexity. One can see just how wrong things have gone when we see that some interpreted sentence types can count as Strawsonian validities without even being true. What use is a characterization of validity which doesn't even secure truth?"

I think there is something to this criticism. The important thing is that it is *as* good a criticism of Strawsonian definitions as it is of their classical counterparts (at least as long as infectious defect is *possible*, as I have been assuming in this chapter—things might be different if this were denied). Classical validity and consequence, to the extent they track a 'true' form of validity on the inferential conception, are MFOF-validity and consequence. And these are also well-behaved only thanks to a stipulative restriction to consider necessarily truth-evaluable sentences. If they were not so stipulated, their character would be completely different (or the notions of validity and consequence unfounded).

And it is true that Strawsonian validity (or consequence) is merely conducive to truth (or truth-preservation), but insufficient for it. Some sentences of an interpreted trivalent language are counted as Strawson validities that are not true. But, of course, this would not happen if the Strawsonian definition were applied only within a language with necessarily bivalent sentences. And, likewise, a classical characterization of validity (that is, MFOF-validity) would classify some untrue sentences as validities, were it applied to a language that contained infectious defect. In this sense, classically valid 'form' is *also* 'merely conducive' to truth, but insufficient for it.

This parity can be disguised by a more or less arbitrary restriction of a consequence relation to *a type of language*—e.g., applying classical consequence only within the strictures of something like a classically interpreted (and so bi-

valent) language. There is a temptation, when focusing on bivalent languages to think that the fact that a sentence has the syntax of a classical validity is 'enough' to secure its truth in this context. But analyzing classical validity into a 'true' form of validity reveals that this is not entirely correct. It is the fact that the sentence ϕ has the syntactic structure of a classical validity *alongside* the fact that it happens to be in a language in which all sentences, including ϕ, are (necessarily) bivalent. The definitions of MFOF-validity (and Strawsonian validity) make the implicit relativization here explicit. And this reveals that anything gerrymandered, or stipulated away, by the Strawsonian definitions is already present in the classical case.

How should we react to this circumstance? Which notion of validity and consequence—MFOF/Strawsonian or wMFOF—is the correct characterization?[12] As far as I can tell, it doesn't really matter as long as one is clear about the implications of using one definition or the other. After all, these definitions are not competing. They each investigate the presence of necessary truth, or truth-preservation, that would be possessed in virtue of different sets of linguistic properties. MFOF-validity tracks the presence of a syntactic form conducive to truth, but insufficient for it absent the semantic property of necessary truth-evaluability. Its associated consequence relation is stronger. wMFOF-validity tracks a syntactic form sufficient for truth, independently of whether a sentence is necessarily truth-evaluable. That virtue is counterbalanced by producing a weaker associated consequence relation. Which of these forms of validity and consequence it makes sense for a theorist to employ should probably depend on their theoretical aims.

I mean for the remarks here to constitute a partial defense the utility of Strawsonian characterizations of validity and consequence in the trivalent context, in part by noting that they are not merely coextensive to classical characterizations but, properly understood and analyzed, effectively identical with them. I suspect that in spite of this equivalence, many theorists will be reluctant to acknowledge a property as a form of validity if that property does not at least secure actual truth. For these theorists, the mere possibility that sentences could bear infectious, inference-blocking defect should drive toward definitions that produce weaker consequence relations like those based on weakened modalized first order form (or whatever properties properly capture the extent of the 'infectious' character of defect—it needn't reflect the weak Kleene pro-

[12] And there are perhaps further variants to consider—see n.10.

jection schemes in particular, for example). But then they should equally be driven to such definitions *even* in the so-called 'classical setting' (at least if that merely means that one is investigating languages where bivalence is ubiquitous). For these theorists there may be a lingering concern about the weakness of the resulting logic. How can we get by with a logic that forbids so many logical inferences? Shouldn't we strive to somehow *avoid* this result? Or, if avoiding it is impossible, isn't the situation lamentable?

Certainly some theorists have recoiled at the weakness of certain logics, and viewed this as something to be avoided. But often enough, little beyond the weakness of the logic is cited as the problem—that is, it is often not stated *why* weakness should be a problem. In the next section, I want to turn to a literature where one ground for concern with weak logics is helpfully articulated. The literature in question concerns the development of formal theories of truth where there is strong pressure to restrict consequence relations in response to the paradoxical behavior of 'liar sentences.' In this context, many theorists have expressed concerns that we have to be cautious in how far we restrict consequence relations, as the more we do so the more we impede *ordinary good reasoning*. On the conception of logic I've been advancing, I think this worry would be borne out of conceptual confusions. Rooting out those confusions can help teach us some general lessons about the significance of weakening a logic in my sense.

9.2 Theories of Truth, Weak Logics, and Ordinary Reasoning

The problem of developing a formal theory of truth is in large measure the problem of finding formally rigorous ways of circumventing the paradoxical properties of liar sentences like (L).

(L) (L) is not true.

If (L) is true, it seems that what it says is true—namely that it is not true. But if (L) is not true, it would seem that what it says is not the case—namely, it is not the case that (L) is not true. So (L) would be true after all.

Familiarly, using plausible principles about truth alongside cherished logical principles in reasoning like the above, one can exploit sentences like (L) to derive a contradiction that appears to rest on no premises. There is no hope of surveying the vast space of formal responses to this worry, let alone their philo-

sophical bases, here.[13] Instead, I want to focus attention on a style of objection which surfaces occasionally in debates over the correct formal theory to adopt, which I will call the "weak logic objection".

The objections first surface in response to the work of Kripke (1975), who developed a family of interpreted languages ('fixed-point constructions') making use of truth-value gaps. A key advantage of Kripke's constructions was that, in them, a truth-predicate could truthfully apply to (names or descriptions of) sentences of that very language, including many sentences which themselves contained a truth-predicate. As Kripke compellingly argued, ordinary intuition seems to support the possibility of such successful self-application of a truth-predicate. Kripke's constructions succeeded in respecting these intuitions while evading the logical problems generated by liar sentences. And they did this by giving paradoxical sentences like (L) a gap value which Kripke glossed as marking failure to express a proposition.[14] The details of Kripke's constructions won't be necessary for our purposes. It will be enough to know that Kripke attempted to capture certain intuitive features of the use of a truth-predicate by making use of a semantic status other than truth and falsity that marked some form of semantic defect.

An explosion of work on truth grew out of reactions to Kripke's proposal. The consensus was that Kripke's constructions represented definite progress but also left important problems unresolved. One key point of dissatisfaction with Kripke's constructions, which I won't touch on directly here, concerns their expressive power. Many theorists claim that Kripke's constructions leave things unexpressed that should be expressible, and for this reason the constructions need to be amended or even scrapped. But another large class of worries were instead directed at the logic of Kripke's constructions—and in particular at the way that this logic departs from a classical ideal. It is these latter objections I want to discuss here.

For example, Gupta and Belnap criticize Kripke's theories for abandoning classical logic, when alternate revision theoretic tools enable us to keep it.[15] Field criticizes fixed-point theories for lacking an 'adequate' conditional—which claim is justified on the basis of the failure of Kripke's conditional to

[13] For an overview, see Beall et al. (2020).

[14] Kripke (1975, 699–700). In drawing on this idea Kripke did hedge that he was "not attempting to be philosophically completely precise."

[15] Gupta & Belnap (1993, 98).

safeguard certain classical validities and inference rules.[16] And these reactions aren't limited to Kripke's work. Beall, for example, also expresses worries about the weakness of Priest's LP, which lacks a conditional that validates Modus Ponens.[17] These are all instances of a very general kind of objection found again and again in the literature on truth: the objection is that the 'weakness' of various logics, in the sense of endorsing few familiar (typically classical) logical entailments, is something that requires strong justification. Absent that justification, stronger logics—especially those approaching classical ones—are to be preferred to weaker ones.

These objections are often given as if it were transparent what the problem with the weak logic is supposed to be. But thankfully there are a few places where objectors elaborate slightly. On these rare occasions where more is said, it is typically alleged that a weak logic poses a threat to *reasoning* (and especially 'ordinary reasoning'). Feferman, for example, voices his concerns with Kripkean frameworks by saying that "…*nothing like sustained ordinary reasoning can be carried out in* [strong Kleene logic]."[18] Field echoes these remarks in his own criticism of the Kripkean framework for lacking an adequate conditional. In the strong Kleene setting, he claims, "…[t]he lack of a[n adequate] conditional (and also of a biconditional) cripples ordinary reasoning."[19] I want to focus on this particular way of framing the problem with weak logics here, since I suspect that a concern with 'saving' ordinary reasoning tacitly underlies many other versions of the objection as well.

Let's take Field's particular presentation of the problem.

> The first [reason that a Kripkean theory based on the strong Kleene scheme is inadequate is] that the theory is too weak to carry out ordinary reasoning. The most notable weakness is that it does not contain a decent conditional or biconditional. One could of course define a conditional from $\neg$ and $\vee$ in the usual classical manner: take $A \supset B$ to be $\neg A \vee B$. (And once you

[16] Field (2008, 72-3). Visser (2004, 204-5) actually argues that all of Gupta and Belnap's criticisms of fixed-point theories ultimately boil down to the criticism that such theories lack of an 'adequate' conditional as well.

[17] Beall (2009, 26).

[18] Feferman (1984, 95), emphasis in the original. Feferman's objection is raised in a context where it is clear the *utility* for deduction in a broadly mathematical context is being prized. It is conceivable that in this context Feferman is less concerned with a 'correct' representation of an independent phenomenon, as opposed to formal tools which could serve ulterior purposes.

[19] Field (2008, 73).

> have a conditional, getting a biconditional is trivial.) But while that does a passable job as a conditional in the presence of excluded middle, it is totally inadequate as a conditional without excluded middle: with $\supset$ as one's candidate for $\rightarrow$, one wouldn't even get such elementary laws of the conditional as $A \rightarrow A$, $A \rightarrow (A \vee B)$, or the inference from $A \rightarrow B$ to $(C \rightarrow A) \rightarrow (C \rightarrow B)$…The lack of a conditional (and also of a biconditional) cripples ordinary reasoning.
>
> (Field, 2008, 72-3)

Let's focus on one of the elementary laws, such as that $A \rightarrow (A \vee B)$. This whole conditional is assigned a gap value in the strong Kleene Kripkean construction when A is assigned a gap value and B is false.[20] So it will not count as a validity, if validity is equated with truth on any reinterpretation of the sentence letters.

But what bearing is this supposed to have on ordinary reasoning? Here is one concern: it seems we can know what is expressed by sentences of the form $A \rightarrow (A \vee B)$ like (9) or (10) a priori, say by inferring those claims from no premises.

(9) If it's raining, then it's raining or it's snowing.

(10) If Sasha is coming, then either Sasha is coming or Waheed is staying at home.

Maybe coming to believe these things without evidence from experience is a key part of our ability to engage in ordinary reasoning. If so, any theory that told us that we could not come to have these beliefs in this way would warp our ability to reason. Perhaps it would prescriptively rule out a legitimate form of reasoning that we otherwise make good use of, or perhaps it would descriptively mischaracterize good reasoning as bad reasoning.

But actually *neither* of these results follows merely from discarding $A \rightarrow (A \vee B)$ as a logical validity. As discussed in §9.1, on conceptions of validity

[20] The strong Kleene projection schemes run as follows:

ϕ	$\neg\phi$
F	T
U	U
T	F

$\phi \wedge \psi$		ψ		
		F	U	T
ϕ	F	F	F	F
	U	F	U	U
	T	F	U	T

$\phi \supset \psi$		ψ		
		F	U	T
ϕ	F	T	T	T
	U	U	U	T
	T	F	U	T

like wMFOF-validity, the fact that something is not logically valid does not mean it cannot be justifiably inferred without premises. What it means is that the necessary truth of what the sentence expresses is not secured *merely* by its possession of a certain set of linguistic properties. It may be that its possession of further linguistic properties does secure its expression of a necessary truth. And as long as that necessity is appreciable, the necessity in question could be inferred without premises, and so known a priori.

In the previous section, I noted that necessary bivalence is a key semantic property ignored by characterizations of validity like wMFOF-validity that can 'pick up the slack' of securing an entailment. Are (9) and (10) necessarily bivalent? Not obviously. But it should probably be stressed that it was equally unclear that we could truly come to know them without some kind of further justification. For example, perhaps one needs to know that both Sasha and Waheed exist (or maybe just that Sasha exists) in order to be in a position to infer (10). The more important point is that any status that (9) and (10) had *before* consideration of paradox is one that they can retain after consideration of paradox, at least as long as they retain their actual truth-conditions. There is no reason to think that the Kripkean theory is in any way committed to changing the truth-conditions of these non-paradoxical sentences. And of course this point about (9) and (10) holds quite generally of any sentence that does not contain a truth-predicate.

Thus, *at least in the simplified form I've presented it*, and given the conception of logic I've been advancing, the concern that invalidating $A \rightarrow (A \vee B)$ on the basis of paradox threatens ordinary reasoning is about as compelling as the concern that classical logic threatens ordinary reasoning by failing to capture the entailment from *o*'s being a vixen to *o*'s being a fox. The objection would stem from a confusion about the relationship between logical entailment and entailment relations more broadly. And there is nothing special about $A \rightarrow (A \vee B)$: we could say the very same of the other validities or inferential forms Field mentions.

Now, there are at least two issues that could complicate the availability of this simply reply.

First, the reply depends on the tenability of a conception of logic like the one I've developed, alongside the presuppositions in the philosophy of language and mind that are necessary to ground it. One of the key presuppositions relevant to the present discussion is that the contents of mental states

have truth-conditional structure, and that we can use this truth-conditional structure to read off a property that is conducive to inferential goodness. On this picture, truth-conditions are prior to, and ground, conditions of inferential goodness that logic explores. But this conception of mental content is controversial, and its reliance on truth-conditions to ground relations of inferential goodness may seem to some theorists to have things backwards. Field, insofar as he advocates for deflationist views that deny truth-conditions a central role in theories of meaning and cognition, may be among those that balk at these foundations.[21] Indeed, in discussing the Kripkean theory based on the strong Kleene scheme, which invalidates the laws he discusses, Field emphasizes that he favors a view on which Kripke's model-theoretic construction is of largely instrumental significance. According to Field, it is a serious mistake to try to identify the notion of truth-in-a-model used by the Kripkean fixed-point construction with the a genuine notion of truth.[22] Once we take this position, I think it is harder to see whether the reply I've made to Field is available.[23] These issues represent deep differences in approaches to the philosophies of language and mind that need to be resolved before logical questions can even get on the table. However those issues shake out, it is significant and I think surprising that the force of the seemingly simple concerns Field raises could end up hinging on these complex questions in the foundations of semantics.

There is a second concern, specific to the setting of paradox, for appealing to the distinction between logical entailments and broader entailment relations in the way that I have been. I've noted that as long as we can ascribe plausible truth-conditions to 'ordinary' sentences, then any entailment relations between them will be safeguarded regardless of how we treat paradox. At the absolute worst, the sentences will simply be relegated to the category of non-logical entailments. Now, paradoxical sentences (insofar as they receive the gap value) won't tend to respect the entailment patterns of these ordinary sentences. But, of course, it is natural to think that paradoxical sentences are far from ordinary. Maybe we do have 'ordinary' ways of reasoning with such sentences. But the idea that a theory was at fault for not safeguarding them

[21] See, notably, Field (1994).

[22] See especially Field (2008, §§3.2,3.4)

[23] As far as I can tell the same form of reply will be available to someone who advocated for, say, an inferential role semantics. They wouldn't lean on truth-conditions to do so, but on a more direct distinction between logical and non-logical inferential roles. But since I find these theories harder to work within, I will stick with my hedge here.

would be an extremely weak objection to the theory. Those ordinary modes of reasoning tend to lead to contradiction—surely it is fair to let some of them go.

It is here that a point familiar from Kripke looms to strengthen the second concern. The point is that paradox can creep into language in unexpected, contingent ways. A reasonable speaker might inscribe the sentence (11), for example, on the whiteboard of room 1029K of the Cathedral of Learning without knowing what room they are in.

(11) The sentence written on the whiteboard of room 1029K of the Cathedral of Learning is not true.

In this way, thinkers with no introspectively detectable incoherence or irrationality can stumble into paradoxical assertion.

Now one might reasonably think that there are 'ordinary' ways of reasoning (and inferring) with the contents expressed by sentences like (11). But we've just seen that these sentences can become paradoxical in ways that we can't always foresee. How can we keep the reasoning that appears to be good (when contingent matters are favorable), while condemning the reasoning that appears bad? Or, worse: shouldn't we say that the reasoning is the *same* regardless of what contingencies arise?

But on reflection contingent paradox doesn't obviously create any substantially new concerns, as the theorist of defect has a natural repertoire of tools to account for reasoning in these contexts. The first thing to note is that the issue raised from paradox here is a perfectly general one arising in the investigation of the relation between semantic defect and reasoning or inference more broadly. For example, speakers seem like they reason with (what is expressed by) sentences with complex demonstratives with unsatisfied nominals, descriptions with unsatisfied descriptive material, and many other forms of sentence that are candidates for treatment with some kind of semantic defect. Each of these kinds of sentence can be 'contingently' defective in unforeseeable ways. I may have excellent (misleading) evidence that makes it perfectly rational to believe that the United States is a monarchy while I utter (1), and I may have equally excellent (misleading) evidence that the animal I am pointing at is a terrier while I utter (3).

(1) # The present king of the United States is in Germany.

(3) # That terrier [pointing at a siamese cat] is dangerous.

If (1) or (3) are defective, then it is clear that one can stumble into asserting actually defective sentence while being perfectly rational. And it seems abundantly clear what accounts for this fact: *in these contexts one is rational in believing or presupposing that circumstances are such that the assertion would be not be defective*.[24]

Once one recognizes this straightforward fact, there is a clear strategy for explaining how we appear to reason well in the presence of certain forms of semantic defect. In the case of propositional expression failure, we could reason with the contents that *would* be expressed by the defective sentences we utter *were* the situation to be as we rationally believe or presuppose them to be. Matters are even more simple in the case of contingently defective trivalent contents. In this case one can reason directly with these contents themselves, but against the backdrop beliefs or presuppositions that the world is such that defect isn't arising.[25] The background assumptions restrict the worlds compatible with one's starting acceptance states to those where a trivalent content has only bivalent values. Inference past that point would be rationally grounded in precisely the ways that bivalent reasoning is—it could behave in a perfectly classical manner, for example.

While there are numerous variations and subtleties to examine,[26] these kinds of simple maneuvers seem to provide resources to cover all instances of good reasoning it is *obvious* that we would have reason to safeguard. The maneuvers won't of course cover cases of explicit reasoning with (say) liar-like sentences that one has adequate evidence are defective. But as I mentioned before, there is no clear grounds for thinking that kind of reasoning is at all important to safeguard to begin with. It certainly can't be billed as 'ordinary' reasoning that needed to be saved.

Perhaps there is something further to be said here on behalf of the weak-logic objector. But it is not clear what it would be. In even the most elaborated cases where weak-logic objections are pressed, the most we find are cursory remarks like those given in the quotations of Field and Feferman above. If there are ways to press the worry further, nothing has been said by purveyors of weak logic objections to make them apparent.

[24] Cf. Shaw (2016, §2.2).

[25] These need not be *linguisticized* beliefs directly about defect. They may amount to little more than the belief or presupposition that the United States has a king, for example.

[26] For example, in cases of defect we are also free to reason *about* the truth of the sentences we express, which provides another store of possible good inferences to work from.

So, the weak-logic objection runs into serious difficulties as soon as we remind ourselves of the importance of the distinction between logical and general entailment relations. As I've flagged, the efficacy of a reply based on that distinction does depend on the characterization of logic I've been appealing to, along with presuppositions in the philosophies of language and mind that went into framing it. But once those are in place, it is not obvious whether any force remains to objection, at least in its present underdeveloped state.

Now, in §9.1 I drew two important lessons for understanding logic in the presence of infectious defect. The first was the importance of respecting the difference between broader entailment relations and logical entailment relations. The second lesson concerned the importance of respecting the relativization of consequence relations to different batches of linguistic properties. So far I've focused on the importance of the first lesson of §9.1 for weak logic objections. But the second lesson has equally important consequences for them.

It is all too often presumed that the logic governing Kripke's theory *is* weaker than classical. For example, we noted above that Gupta and Belnap object to the Kripkean system for abandoning classical logic. This is understandable. Typically the consequence relation in the trivalent setting is defined in terms of truth preservation across all trivalent models of a certain type. Though I didn't discuss it explicitly, we could of course formulate an analog of weakened modalized first-order form that respected the projection behavior of the strong Kleene scheme (rather than the weak Kleene scheme), and use it to define a consequence relation which has the relevant features. A lesson of §9.1, of course, was that if we define the consequence relation in this way, then its weaknesses will also be apparent in the 'classical' purely bivalent setting. That is, there would be nothing 'weak' introduced by the Kripkean system, that *wasn't already present in that bivalent setting*. Accordingly, it would be hard to know how one could object to the Kripkean system on the basis of 'its' logic.

Conversely, I noted in §9.1 that we are free to use FMOF-validity and consequence (equivalent to the Strawsonian definitions) in the setting of semantic defect. The result of doing so, I noted, was classical logic. I argued that if it is *ever* legitimate to view the classical relations as a form of consequence, the treatment of sentences in the languages of defect with those relations was equally legitimate. Accordingly there is a perfectly legitimate sense in which Kripke's system constituted no change in logic whatsoever. Indeed, regardless of how

we define our consequence relation, this holds true.

Maintaining that Kripke's system is classical may seem radical. But Kripke himself seems to have anticipated this idea (and indeed expressed surprise that philosophers would interpret him any other way). In a telling footnote, Kripke has the following to say.

> I have been amazed to hear my use of the Kleene valuation compared occasionally to the proposals of those who favor abandoning standard logic "for quantum mechanics" or positing extra truth values beyond truth and falsity, etc. Such a reaction surprised me as much as it would presumably surprise Kleene, who intended (as I do here) to write a work of standard mathematical results, provable in conventional mathematics. "Undefined" is not an extra truth value, anymore than—in Kleene's book—u is an extra *number* in sec. 63. Nor should it be said that "classical logic" does not generally hold, any more than (in Kleene) the use of partially defined functions invalidates the commutative law of addition. *If* certain sentences express propositions, any tautological truth function of them expresses a true proposition. Of course formulas, even with the forms of tautologies, which have components that do not express propositions may have truth functions that do not express propositions either…Mere conventions for handling terms that do not designate numbers should not be called changes in arithmetic; conventions for handling sentences that do not express propositions are not in any philosophically significant sense "changes in logic." The term 'three-valued logic', occasionally used here, should not mislead. All our considerations can be formalized in a classical metalanguage.
>
> Kripke (1975, 700-1, n.18)

Here Kripke leans on the interpretation of his gap value as marking failure to express a proposition, and uses it to claim that his system does not represent a change in logic. Indeed, Kripke maintains that nothing prevents us from saying that classical logic holds generally—even appealing to the now-familiar technique of Strawsonian relativization ("*If* certain sentences express propositions …") to show this. But how are these claims justified?

Kripke's analogy draws on a a critical, tacit assumption about logical inquiry. Kripke notes that in Kleene's system, the value 'u' is not another num-

ber, but a bookkeeping device to keep track of terms failing to designate numbers. Things like the commutative law of addition are generalizations about numbers. And, of course, merely noting that we can sometimes use a term that fails to speak of a number shouldn't influence our theory of how the numbers behave. Kripke claims his case is analogous, noting that his gap value is tracking a similar kind of expressive failure to Kleene's. In Kripke's case, the third value marks a failure to express a proposition, not a new sort of semantic status for a proposition to bear.

The analogy with Kleene's case would be complete if we could just claim one more thing: that logic is most fundamentally concerned with the behavior of the *contents expressed by sentences* (rather than the interpreted sentences themselves) in just the way that the theory of partial functions is most fundamentally concerned with the behavior of the numbers (rather than our interpreted formalism for talking about them). For if that were true, we could say that what logic was most fundamentally investigating remains unchanged on Kripke's proposal, as we merely explore new, fallible modes of speech about it.

Kripke's tacit assumption here—that logic has a fundamental concern with properties of assertoric content—is one of the key components of the conception of logical consequence that I've defended. Granted, I take logic to have this concern because logic is further concerned with (indirectly) investigating properties of such contents that contribute to inferential goodness. I doubt that Kripke had that specific justification for the concern with properties of assertoric content in mind. What is important is that there is at least one elaboration of a conception of logic that would make perfect sense of the remarks Kripke makes here.

When Kripke says that we can allow that classical logic holds generally in his framework, there are two ways of maintaining this claim in light of what Kripke says. On a weaker construal, Kripke is only claiming that no change of substance has been proposed for the behavior of the entities of fundamental concern for logical inquiry, and in *this* sense nothing has changed from the classical setting, even though the details of the logic of his system will not actually vindicate classical inference relations. Alternatively, Kripke could have intended a stronger claim: that when we define the consequence relation for his language, the result should actually result in classical entailment relations.

Whichever claim Kripke intended, both are perfectly justifiable. As we've noted, MFOF-validity and consequence are applicable to the Kripkean setting

to produce completely classical results. And the application of this form of validity and consequence is just as legitimate in the Kripkean setting as it ever was in the bivalent setting. If *anything ever* 'obeys' classical logic, Kripke's system does as well, and in the very same sense.

Recall that my first reply to weak logic objections on the basis of the distinction between general and logical entailment relations depended on the viability of my conception of logic. The same holds of this new reply. More particularly, we need to justify some conception of logic on which it is *fundamentally* concerned with assertoric contents, so as to vindicate Kripke's key tacit assumption.[27]

But, again, none of these caveats should detract from the point that objections to the logic of Kripke's system are strikingly underdeveloped. Not only is it unclear whether the alleged weakness of Kripke's system is problematic, but it is unclear whether there is any weakness in the system to begin with. Purveyors of weak logic objections take both the presence of weakness and its objectionability largely for granted. This seems especially problematic with respect to Kripke, who articulated a reasonable basis for thinking that his proposal represented no departure from the classical setting. The view of logic that I've provided supplies one way of fleshing out, and further grounding, Kripke's contention. Perhaps there are flaws in the picture I've advanced, or the reasoning I've given from it. But one thing is clear: weak logic objections have no force insofar as they rest on the underdeveloped, merely intuitive grounds on which they've relied so far.

9.3 Reflections

I want to take a brief moment to step back and draw a general lesson about the complexities that semantic defect introduces into our thinking about logical consequence, at least within my proposed framework.

[27] One may wonder whether the reply here also depends on Kripke's controversial characterization of paradox as resulting in propositional expression failure rather than, for example, the expression of contingently defective propositions. For the record, I think that when the details of contingent paradox are looked into, it is very hard to maintain Kripke's treatment of paradox as mere failure to express a proposition, rather than defect possessed by trivalent content (see Shaw (2021/ms.) and the citations therein for arguments to that effect). Still, as discussed in §9.1, there are interpretations of trivalent defect which will again legitimate the Strawsonian move that safeguards classical logic in this setting. So I don't see Kripke's more controversial stance on the defect of paradox as integral to the reply.

What we've uncovered is that the *mere possibility* of semantic defect makes good inference harder to track, merely by looking at very minimal sets of semantic properties which, notably, don't include the presumptive property of truth-evaluability. This creates a theoretical choice point. On the one hand, one could try to capture information about a broader range of entailments by looking at formal features of sentences types that are conducive to, but (even bracketing appreciability) insufficient for, good inference. On the other, one could investigate formal features (bracketing appreciability) sufficient for good inference, but at the cost of relegating very many entailment forms to 'non-logical' status. The choice between these two paths is reflected in the MFOF and wMFOF characterizations of validity and consequence respectively.

If there is any 'problem' raised by semantic defect in any of this, it is a problem *for the hopes of the theorist*. The theorist may have hoped that the object of their investigation—a property of linguistic content conducive to good inference—was easily recognizable merely from the weakly specified 'forms' of the sentences that express that content. If infectious forms of semantic defect like those sketched in §9.1 are even possible, this is simply not true. As such, the starting hopes of the theorist are dashed.

But it is not clear that there is any further problem, and absent substantial argument certainly none that calls for attempts to shore up or improve our logical theories by strengthening them, or making them better behaved. In particular, there is no clear problem for reasoners, the goodness of whose inferences we aim to characterize. No uncontested good inference by ordinary reasoners is threatened by the presence of semantic defect, since no uncontested good inference itself involves the presence of semantic defect.[28] The idea that defective contents (or lack of contents) could influence the goodness of reasoning with non-defective contents is an illusion created by the core logical idea of investigating relationships among contents while abstracting from some of their semantic features. Of course, if we abstract from whether defect is present, good inferences will share a 'form' with bad ones. But that is just a result of the process of abstraction—one that speakers obviously do not (or at least need not) replicate in their own reasoning.

The lesson might be summed up as follows: semantic defect represents

[28] Or, more precisely (to acknowledge apparent good inference in the presence of contingent defect): no uncontested good inference involves the rationally recognized presence of semantic defect.

an obstacle for modelers of reasoning, not an obstacle for reasoners modeled. The idea that a problem for the theorist of reasoning is a problem for the reasoner arises from overestimating the importance of logical theorizing. Logic's study of reasoning is limited in so many ways: logic studies inference, which is but one of several components of reasoning; logic studies a necessary, but insufficient condition on good inference; and logic investigates good inference ascertainable from language while abstracting from some linguistic properties. Each of these limitations opens up space for good reasoning to proceed in ways which logic cannot 'see.' But that is not necessarily any fault in the logic.

The point of logic was precisely to investigate those aspects of inference which are schematically repeatable, easily isolatable, and formalizable. When there is any increase in the complexity of the relations between language and the contents that underlie good inference, that tends to correspondingly tighten the boundaries on that well-behaved subset of formalizable good inferences. But recognizing a tightening of the boundary *within* the class of good inferences never affects the class of good inferences more broadly, nor the nature of good reasoning. It is only when we presuppose that all reasoning must have the properties suiting it to logical study that a weakness present in a logical system begins to look like a theoretical fault. But far from being such a fault, the weakness appears to simply be an adequate reflection of complexity in the phenomenon that logic aims to study.

CHAPTER 10

Validity in the Presence of Perspectival Thought, Context-Sensitivity, and Ambiguity

In this chapter, I examine how inferential logical investigation would be affected by three distinct but interrelated phenomena:

- *perspectival thought* (i.e. thought about the how the world is or might be from a particular perspective within it);
- *context-sensitivity* (i.e. the phenomenon in which the semantic contribution of a word conventionally varies with its linguistic context of use); and
- *lexical ambiguity* (i.e., the phenomenon in which a single orthographic or phonological type creates divergent semantic contributions tied to different linguistic conventions).

I will be guided in my discussion of all three of these topics by an engagement with the seminal work of Kaplan (1989b,a) on the semantics and logic for perspectival context-sensitive terms like "I", "here", "now", "actually", "this", "that", "there", and deitic uses of personal pronouns like "he"/"she"/"it". These are terms whose reference shifts from context of use to context of use (hence their context-sensitivity). But they are also terms whose proper usage intuitively accompanies perspectival thoughts (hence the qualifier that these are specifically *perspectival* context-sensitive terms). The goal will be to learn from Kaplan's insights, but also ultimately to critique some aspects of his formalism and its philosophical basis.

I begin with a largely expository section in §10.1 reviewing the semantic and logical underpinnings of Kaplan's logical system LD for perspectival context-sensitivity. I highlight a curious existence entailment within Kaplan's system,

and also discuss an apparent challenge the system presents for thinking that logical truths are metaphysically necessary (closely connected to that discussed in Chapter 8). I note that Kaplan is helpfully explicit about his understanding of logical validity, and frames it in terms disconnected from inference.

Reflecting on Kaplan's system will lead us, in §10.2, to consider distinct but related non-linguistic questions about the nature of perspectival thought. As is familiar from the literature on *de se* cognition, philosophical puzzles about perspectival thought can be raised and addressed independently of linguistic considerations. I review the underpinnings of contemporary 'exceptionalist' views of *de se* cognition, according to which perspectival thought calls for special accommodation in our theories of attitudes. I then explore the natural logic that would model deductive inference for such exceptionalist perspectival thoughts. The result is a system LD^* which closely resembles Kaplan's framework, but with several important differences. Notably LD^* invalidates the suspicious existence entailment of LD. And LD^*'s philosophical underpinnings actually forge strong conceptual connections between logical validity and the generalization of metaphysical necessity for *de se* information.

After a brief discussion of how to accommodate the passage of time for inferences involving perspectival thought in §10.3, I turn back to consideration of language. For heuristic and comparative purposes, I explore how a deductive inferential logic should respond to the integration of 'unresolved' lexically ambiguous terms in §10.4. I note that while we can develop formal systems that forego access to disambiguating information, the resulting systems tend to undergenerate in modeling good deductive inferences in several predictable ways. I argue they also suffer from a more general problem of making a 'logic' into a study *of* language (in particular of the contingencies of orthography or phonology) in its relation to inference, rather than a study of inference *through* language. As a result, these 'logics' cease to be usable to model good *bases* for deductive inferences.

In §10.5 I explain why, in a logic tracking good deductive inference, one would anticipate some parallels between a logical treatment of linguistic context-sensitivity and ambiguity. Following these parallels up, I argue that while many of Kaplan's modeling choices in developing LD would be obligatory for an inferential logic for perspectival thought they are far from obligatory, and even in some ways idiosyncratic, in developing an *inferential* logic for context-sensitive expressions—even for the perspectival context-sensitive

terms on which Kaplan focused. I explain why Kaplan's logic can intuitively undergenerate in modeling good deductive inference in ways similar to logics for ambiguous terms that forego access to disambiguating information. I argue that we can see modifications that Kaplan explored for his system when accommodating true demonstratives (like "that") as resulting from a particularly problematic instance of this deficiency. What is more, as soon as we consider *non*-perspectival context-sensitive terms (like gradable adjectives, quantifiers, and modals), incorporating Kaplan's structural choices in developing a inferential logic of context-sensitivity would become highly arbitrary, if not obfuscating. Tellingly, in the setting of non-perspectival context-sensitivity it also becomes clear that a logic helping to track the goodness of deductive inferential relations again must maintain strong conceptual ties between validity and metaphysical necessity.

The main conclusions of this chapter concern the development of an *inferential* logic for context-sensitivity and perspectival thought that, it should be borne in mind, Kaplan did not pursue. Even so, in §10.6, I step back to review the lessons of the chapter and use them to back up a conjecture: that Kaplan's own logic for context-sensitive terms reflects a periodic conflation of linguistic context-sensitivity and perspectival thought with the result that Kaplan's system, and the philosophical basis underlying it, are a kind of hybrid that fails to faithfully model either phenomenon. (This charge, it is worth stressing, is compatible with Kaplan's *compositional semantics* for context-sensitive expressions being fully accurate.) I acknowledge that fully defending this conjecture requires investigation that goes beyond the scope of this book, and conclude by highlighting the key tasks that remain.

10.1 Kaplan's Logic of Demonstratives

Some expressions, like "I", "here", "now", "this", and "that" change their denotations from use to use. Kaplan (1989b) dubbed such expressions *indexicals*, provided the standard compositional framework for them, embedded it within a broader philosophical view about their meanings, and formulated a notion of logical validity that could apply to sentences containing context-sensitive terms.

Kaplan distinguished between *contexts of utterance* and *circumstances of evaluation*. The former are roughly those settings under which a sentence

could be uttered, and the latter are roughly those settings relative to which the truth of some claim can be ascertained. On the Kaplanian picture, indexicals require us to distinguish between two kinds of meaning that interact differentially but systematically with contexts of utterance and circumstances of evaluation. First, each simple expression of a language—and by extension each composite expression—can be used to express what Kaplan calls *contents*. Kaplan took this to correspond at the level of the whole sentence to 'what is said' with our words, and had such contents play many of the roles traditionally reserved for linguistic propositional content. But Kaplan suggested that indexicals require us to also accommodate a second tier of meaning called *character*, which serves as something like a rule for generating content as a function of an expression's context of use. Character is what enables us to exploit features of our surroundings in systematic ways to facilitate the expression of certain contents.

For example, if I say "I am hungry" and you reply by saying "you are hungry, and I am hungry too," then our utterances of "I am hungry" share a Kaplanian character. (The words used in the sentences correspond, say, to the same dictionary entries.) But they differ in content. My words report *my* hunger, your words reports *yours*—so our both using them does not reflect our agreement on one fact, but our statement of different facts. By contrast, my utterance of "I am hungry" and yours of "you are hungry" differ in character, as "I" and "you" are associated with different rules for selecting a referent—one on the basis of who is speaking, the other on the basis of who is being addressed. Even so, the content of our utterances is the same: we are both claiming, of me, that I am hungry.

Formally, for Kaplan, a "context is a package of whatever parameters are needed to determine the referent, and thus the content, of the directly referential expressions of the language." (Kaplan, 1989a, 591) He represented it as a tuple of an agent, time, place, and world. To focus on tuples that actually represent a possible setting for an utterance, he defined a *proper context* as one at which the agent of the context exists at the time and place of the context, in the world of the context. A circumstance of evaluation is modeled as a set of parameters relative to which what is said can be evaluated, and includes features like a time, a place, and a world. Kaplan noted that time- and world-shifting operators (like "in 10 months", "it is metaphysically possible that") not only shift circumstances of evaluation away from a given actual context of utter-

ance, but may shift it away from *any* proper context. We sometimes have reason to describe what things would be like where no possible speakings could transpire. So there are no special constraints built into the relationships among the parameters of a circumstance of evaluation.

Semantics must recursively and simultaneously keep track of the influence of both contexts and circumstances of evaluation on extension assignments, because if an indexical is embedded within an intensional operator that shifts the circumstance of evaluation, then the indexical still receives its extension assignment from the context of utterance. For example, in the sentence "in 12 months, those now alive are dead", we evaluate the embedded "those now alive are dead" relative to a circumstance of evaluation shifted 12 months forward in time from the time of the context of utterance. But in spite of the shift, the expression "now", and so also "those now alive", nevertheless 'reaches back' to the original time of the context of utterance for its interpretation. The sentence thus asks us to look at the set of persons alive at the time of the context of utterance, and then claims of them that they have perished by the shifted time of evaluation twelve months in the future. Accordingly, the extension of the embedded clause "those now alive are dead" must be evaluable with respect to *two* times: the time of the unshifted context, and the time of the shifted circumstances of evaluation.[1]

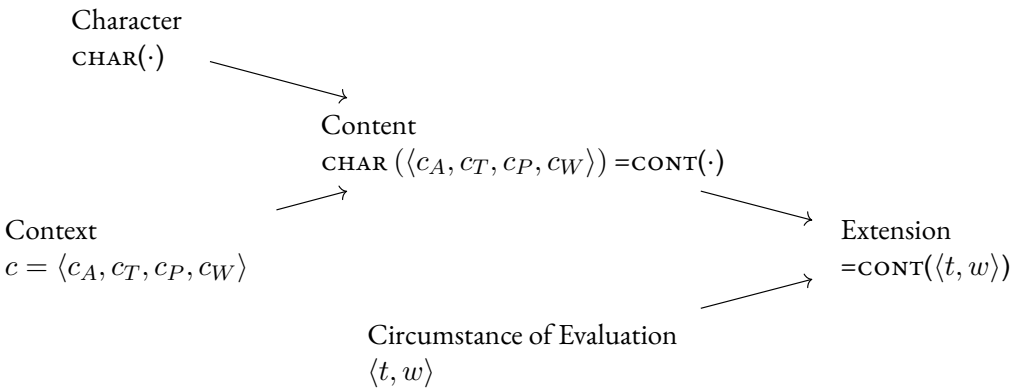
Character (formally given as a function CHAR) determines content as a function of context (formally representable as a tuple of an agent, time, place, and world $\langle c_A, c_T, c_P, c_W \rangle$). And content (formally given as a function CONT)) determines extension as a function of circumstances of evaluation (formally given as a time and world $\langle t, w \rangle$). This leads to a familiar 'two tier' determination of extension seen in Figure 10.1.

Kaplan formalized the recursive tracking of character and content in a logic he called LD, which he also used to characterize a notion of validity for sentences containing indexicals (where those indexicals were treated as logical vocabulary). Here I want to outline a very slightly modified *fragment* of the logic LD developed by Kaplan, which I will call LD^-, since this fragment will suffice to raise the logical issues I want to discuss.[2] The language of LD^- is given

[1] This phenomenon, sometimes called 'double-indexing', was already appreciated in the case of time by KAMP (1971).

[2] Some important differences: Kaplan's LD is two-sorted (for individuals and places), contains more tense and modal operators, a description operator "the", and a rigidifying operator "dthat".

FIGURE 10.1: Character and Content



by augmenting a first-order language containing identity (but without function symbols) with the following resources:

- A modal operator $\Box$ (it is necessary that)
- Tense operators F,P (it will be/has been the case that)
- A tense operator N (it's now the case that)
- An individual constant I
- A positional constant HERE
- A one-place predicate EXIST
- A two-place predicate LOCATED (is located at)

Terms, formulas, and sentences are recursively defined in the familiar way. A semantics is given through a structure in the following sense.

An LD^- structure is a tuple $\mathfrak{A} = \langle \mathcal{C}, \mathcal{W}, \mathcal{U}, \mathcal{P}, \mathcal{T}, \mathcal{I} \rangle$ such that:

(I) $\mathcal{C}$ is a nonempty set of contexts $c = \langle c_A, c_T, c_P, c_W \rangle$ with:
  (i) $c_A \in \mathcal{U}$ (the agent of c),
  (ii) $c_T \in \mathcal{T}$ (the time of c),
  (iii) $c_P \in \mathcal{P}$ (the position of c), and
  (iv) $c_W \in \mathcal{W}$ (the world of c);

(II) $\mathcal{W}$ is a nonempty set (of worlds);

(III) $\mathcal{U}$ is a nonempty set (of individuals);

(IV) $\mathcal{P}$ is a non-empty set (of positions, common to all worlds) with $\mathcal{P} \subseteq \mathcal{U}$;

(V) $\mathcal{T}$ is the set of integers (thought of as times, common to all worlds);

(VI) $\mathcal{I}$ is an interpretation function such that for each world $w \in \mathcal{W}$ and time $t \in \mathcal{T}$…

(i) for each n-ary predicate P, $\mathcal{I}(w, t, P) \subseteq \mathcal{U}^n$, and

(ii) for each constant symbol a, $\mathcal{I}(w, t, a) \in \mathcal{U} \cup \{\dagger\}$ (where $\dagger$ is some element not in $\mathcal{U}$);

(VII) $i \in \mathcal{U}$ iff $\exists t \in \mathcal{T}, \exists w \in \mathcal{W} : \langle i \rangle \in \mathcal{I}(w, t, \textsc{exist})$;

(VIII) If $c \in C$ then $\langle c_A, c_P \rangle \in \mathcal{I}(c_W, c_T, \textsc{located})$;

(IX) If $\langle i, p \rangle \in \mathcal{I}(w, t, \textsc{located})$ then $i \in \mathcal{I}(w, t, \textsc{exist})$.

Clause (VIII) encapsulates the idea that at a 'proper' context, the agent is at the time, place, and world of the context. Then (IX) guarantees the agent *exists* at the context, by claiming that being at a location at a time in a world guarantees existence.

A (variable) assignment f is a function from variables of the language to elements of the domain $\mathcal{U}$, and f^{α}_{x} is the function that differs from f at most in that it assigns variable α to x. We define the denotation of a term as usual, excepting that I and HERE always denote the agent and place of the context respectively.

> $|\alpha|^{\mathfrak{A},f}_{c,t,w}$, or the *denotation of a term* α *in structure* $\mathfrak{A}$*, with respect to assignment* f*, context,* c*, time* t*, and world* w, is defined as follows:
>
> (I) if α is a variable, $|\alpha|^{\mathfrak{A},f}_{c,t,w} = f(\alpha)$,
>
> (II) if α is a constant other than I or HERE, $|\alpha|^{\mathfrak{A},f}_{c,t,w} = \mathcal{I}(w, t, \alpha)$,
>
> (III) $|\mathrm{I}|^{\mathfrak{A},f}_{c,t,w} = c_A$,
>
> (IV) $|\textsc{here}|^{\mathfrak{A},f}_{c,t,w} = c_P$,

And (assuming the language only contains negation and conjunction as connectives, and universal quantification), we define truth in a structure at a context and circumstance as follows.

$\models^{\mathfrak{A},f}_{c,t,w} \phi$, or ϕ is *true* in structure $\mathfrak{A}$, with respect to assignment f, context, c, time t, and world w, is defined as follows:

(I) $\models^{\mathfrak{A},f}_{c,t,w} P(\alpha_1 \ldots, \alpha_n)$ iff $\langle |\alpha_1|^{\mathfrak{A},f}_{c,t,w}, \ldots, |\alpha_n|^{\mathfrak{A},f}_{c,t,w}\rangle \in \mathcal{I}(w,t,P)$

(II) $\models^{\mathfrak{A},f}_{c,t,w} \phi \wedge \psi$ iff $\models^{\mathfrak{A},f}_{c,t,w} \phi$ and $\models^{\mathfrak{A},f}_{c,t,w} \psi$.

(III) $\models^{\mathfrak{A},f}_{c,t,w} \neg\phi$ iff $\not\models^{\mathfrak{A},f}_{c,t,w} \phi$

(IV) $\models^{\mathfrak{A},f}_{c,t,w} \alpha = \beta$ iff $|\alpha|^{\mathfrak{A},f}_{c,t,w} = |\beta|^{\mathfrak{A},f}_{c,t,w} \neq \dagger$[3]

(V) $\models^{\mathfrak{A},f}_{c,t,w} \forall\alpha\phi$ iff for all $i \in \mathcal{U}$: $\models^{\mathfrak{A},f^{\alpha}_{i}}_{c,t,w} \phi$.

(VI) $\models^{\mathfrak{A},f}_{c,t,w} \Box\phi$ iff $\forall w' \in \mathcal{W}$, $\models^{\mathfrak{A},f}_{c,t,w'} \phi$

(VII) $\models^{\mathfrak{A},f}_{c,t,w} F\phi$ iff $\exists t' \in \mathcal{T}$ with $t' > t$ and $\models^{\mathfrak{A},f}_{c,t',w} \phi$

(VIII) $\models^{\mathfrak{A},f}_{c,t,w} P\phi$ iff $\exists t' \in \mathcal{T}$ with $t' < t$ and $\models^{\mathfrak{A},f}_{c,t',w} \phi$

(IX) $\models^{\mathfrak{A},f}_{c,t,w} N\phi$ iff $\models^{\mathfrak{A},f}_{c,c_T,w} \phi$

Finally we define 'truth in a context' (roughly, truth of what a sentence would say in a context of utterance), logical truth, and logical consequence as follows.

ϕ is true in a context c (relative to LD^- structure $\mathfrak{A}$) iff for every assignment f, $\models^{\mathfrak{A},f}_{c,c_T,c_W} \phi$

ϕ is a LD^- logical truth, $\models_{LD^-} \phi$, if for every LD^- structure $\mathfrak{A}$ and context c of $\mathfrak{A}$, ϕ is true in context c relative to $\mathfrak{A}$.

ϕ is a logical consequence of Γ, $\Gamma \models_{LD^-} \phi$, if for every LD^- structure $\mathfrak{A}$ and context c of $\mathfrak{A}$, if all members of Γ are true in context c relative to $\mathfrak{A}$, then ϕ is true in context c relative to $\mathfrak{A}$.[4]

Here are some familiar validities and consequences that emerge within LD^- (and Kaplan's original LD as well):

(1) $\models_{LD^-}$ EXIST(I)

I exist.

[3] The restriction of true identities to denoting terms is a departure from Kaplan's definition. It won't make much of a difference until we consider some modifications in §10.2.

[4] Kaplan only defines validity for his language, but the extension to consequence is natural and straightforward.

(2) $\models_{LD^-} N\textsc{located}(\text{I}, \textsc{here})$

I am presently located here.

(3) $\text{I} = James \models_{LD^-} N\textsc{located}(James, \textsc{here})$

I am James ∴ James is now located here.

(4) $\textsc{exist}(\text{I}) \wedge \textsc{here} = Pittsburgh \models_{LD^-}$
$N\textsc{located}(\text{I}, Pittsburgh)$

I exist and Pittsburgh is here ∴ I am now located in Pittsburgh.

"I exist" is valid because every proper context is one in which the agent of the context exists. "I am presently located here" is valid because each proper context is one in which the agent of the context is located at the place of the context, at the time of the context. "I am James" entails "James is now located here" because the first claim is only true at contexts where James is the agent of the context, and at all such contexts that agent is again located at the place of the context at the time of the context. Similarly, "I exist and Pittsburgh is here" is only true at contexts where the location of the context is Pittsburgh. At all such contexts the agent of the context is located in Pittsburgh.[5]

The rule of necessitation fails for validities like the first two above. That is.

(5) $\not\models_{LD^-} \Box\textsc{exist}(\text{I})$

Necessarily, I exist.

(6) $\not\models_{LD^-} \Box N\textsc{located}(\text{I}, \textsc{here})$

Necessarily, I am presently located here.

"Necessarily, I exist" is not valid, because even if the agent of the context exists at the world of the context, they need not exist at all possible worlds (which is what the necessitation would require). And "necessarily, I am presently located here" is not valid because the agent need not be located at the place of the context at all possible worlds. LD^- consequences also do not correspond to necessarily true conditionals (even if we were to add the requirement that names like "James" and "Pittsburgh" are rigid).

[5] The premise "I exist" is not needed in LD or LD^- to ensure the entailment. But there is an open question about whether it would count as a good deductive inference without it. This will become clearer in §10.2.

(7) $\not\models_{LD^-} \Box(\text{I} = James \rightarrow N\textsc{located}(James, \textsc{here}))$

Necessarily, if I am James, James is now located here.

(8) $\not\models_{LD^-} \Box(\textsc{exist}(\text{I}) \wedge \textsc{here} = Pittsburgh \rightarrow N\textsc{located}(\text{I}, Pittsburgh))$

Necessarily, if I exist and Pittsburgh is here, then I am now located in Pittsburgh.

The antecedent of the embedded conditional in (7) requires that the speaker of the *context* be whoever James is at the world of evaluation shifted by the modal, which can easily be fulfilled even if James is not 'now' present at the place of the context in that world of evaluation, which is what the embedded consequent requires. The antecedent of the embedded conditional in (8) requires the place of the context to be wherever Pittsburgh is at the shifted world of evaluation. This is easily satisfied even if the speaker of the context is not 'now' present in Pittsburgh in that given world of evaluation.

In effect, all these failures in (5)–(8) arise because circumstances of evaluation can be shifted away from proper contexts. It is proper contexts that determine whether a sentence is valid. But when we embed a sentence that would be true at all proper contexts *under* a modal operator, that operator may evaluate the sentence relative to shifted circumstances of evaluation that correspond to no such proper context. The conditions imposed by the sentence may thus fail to hold in those shifted circumstances, even if they hold in all proper contexts.

The validities in (1)–(4) along with the corresponding failures of necessitation or 'necessary truth-preservation' encapsulate the distinctive features of LD^- (shared by Kaplan's original LD).[6] Kaplan reports that this "convincingly deviant" modal logic is one of the things about LD that he finds attractive.[7]

The failure of a rule of necessitation can appear to reveal that there are contingent logical truths. Kaplan, to my knowledge, never goes quite so far as to say this. In fact, the idea that there could be contingent *or* necessary logical

[6] There is also a source of failures of necessitation in Kaplan's LD that intriguingly arises *even* for some validities not containing indexicals, but which disappears in LD^-. Kaplan's LD types individuals and locations differently, which allows for the expression "some (individual) exists" that is true at all proper contexts, but whose necessitation can fail if there are worlds where no individual exists at those worlds.

[7] Kaplan (1989a, 593).

truths appears to be a confusion within Kaplan's system. Consider Kaplan's remark:

> How can something be both logically true, and thus *certain*, and *contingent* at the same time? In the case of indexicals the answer is easy to see:
>
> E. Corrollary 3. *The bearers of logical truth and contingency are different entities. It is the character (or the sentence, if you prefer) that is logically true, producing a true content in every context. But it is the content (the proposition, if you will) that is contingent or necessary.*
>
> (Kaplan, 1989b, 539)

Note that the 'answer' to the question that opens the above quotation is in fact a rejection of the question. How can something—*one* thing—be both logically true and contingent? The answer is that this is not so in the logic developed. For there are two things, one of which is logically true, the other of which is contingent. In his "Afterthoughts" Kaplan reiterates these ideas.

> I find it useful to think of validity and necessity as *never* applying to the same entity. Keeping in mind that an actual-world is simply the circumstance of a context of use, consider the distinction between:
>
> (V) No matter what the context were, ϕ would express a truth in the circumstances of that context
>
> and:
>
> (N) The content that ϕ expresses in a given context would be true no matter what the circumstances were.
>
> The former states a property of sentences (or perhaps characters): validity; the latter states a property of the content of a sentence (a proposition): necessity.
>
> (Kaplan, 1989a, 596)

The only clear sense in which Kaplan explicitly claims that some logical truths are not necessary is one in which he likewise maintains that *all* logical truths are not necessary (and also not contingent): the logic of demonstratives reveals

for him that logical truths are *not the sort of thing* to be necessary or contingent to begin with. Kaplan need not have spoken in this way, but he did.[8]

Still, to divorce the bearers of logical truth from the bearers of necessity is to deny logical truths are necessary in *some* sense. So whether or not we follow Kaplan's usage, one can reasonably worry that the logical framework poses a challenge to an understanding of logic as tracking relations of necessary truth preservation, as I have proposed we do. And indeed many have taken Kaplan's framework to have shown this. Zalta, whose views were discussed in Chapter 8, seems to think that, up to concerns about the logicality of indexicals, Kaplan's work reveals that logical truths could be metaphysically contingent. More recently, Gillian Russell takes Kaplan's framework to help reveal that characterizing logical consequence in terms of necessary truth-preservation on *any* notion of necessity involves a mistake.

> A standard, if informal, definition of logical consequence is as follows: a sentence A is a logical consequence of a set of premises Γ if and only if *it is impossible for all the members of* Γ *to be true and* A *false*. *LD* shows us that this is a mistake, by which I mean not just that it is a somewhat imprecise characterization that requires more scholarly formal explication, but that it is a step in the wrong direction.
>
> (Russell, 2012, 198)

Just as with the threat posed by an actuality operator in Chapter 8, I am dubious that Kaplan's work on indexicals supports claims like this. Understanding why will, as before, require getting clearer on the relationships between the

[8] One salient alternative would be to treat validity as grounded in a *relation* between content and character. We can say a content is valid* *relative to a character in some context* just in case the content is determined by the character and the context, and the character is logically valid in Kaplan's sense. Finally we could say a content is valid** (simpliciter) just in case it is valid* relative to some character in some context. Compare the corresponding move for the guise theorist, in Salmon (1993), of defining *apriority* (which is also a property of character for Kaplan) as a property of content which it bears in virtue of more fundamental relations between contents and guises under which content can be entertained. Of course, this redefinition comes with unusual consequences familiar from the guise-theoretic case. For example, "James is in Pittsburgh on October 17th 2020"—and indeed any claim about where an agent is at some time, true at some world—could come out valid** for expressing a content that can equally be expressed by "I am here now" relative to the relevant context. Indeed, there is a further worry (an extension of an analogous worry for the guise theorist) that all contents will come out valid**.

proposed compositional apparatus, the assertoric contents expressed by sentences of the system, and the mental contents that would correspond to those assertoric contents. But matters will be much more complicated than in Chapter 8, not least because *unlike* with Zalta and Hanson, Kaplan *did* speak quite directly to some of these issues and produce arguments about them.

We will get to these remarks of Kaplan soon. Before we do, though, I want to highlight three points.

(A) Some validities and consequences within $LD^{(-)}$ seem transparently problematic as ways of modeling good deductive inferences.

(B) Claim (A) is not in tension with any claim Kaplan defends in his work (of which I am aware). In fact, Kaplan is admirably clear about what validity *is* for him, and it is *not* characterized in terms of deductive inference nor anything which has straightforward ties to it.

(C) In spite of this, *many* validities and consequences within $LD^{(-)}$ (including some which contribute to the violation of the rule of necessitation) *do* seem to correspond to good deductive inferences distinctive of reasoning naturally expressed using indexicals.

Let met start with the first point about problematic inference. The most troubling cases in LD for its interpretation as tracking good deductive inference are various existence entailments like (1) and the implicit existence claim in (2).

(1) $\models_{LD^-} \textsc{exist}(\mathrm{I})$

I exist.

(2) $\models_{LD^-} \mathcal{N}\textsc{located}(\mathrm{I}, \textsc{here})$

I am presently located here.

Kaplan jokes that Descartes' *Cogito* ("I think, therefore I am") had one premise too many.[9] I think it is significant this is cast as a joke. It is striking to think that the entire modern historical tradition simply overlooked a superfluous premise for what should have been an elementary inference. Intuitively, it is a logical leap to infer one's own existence from no premises. And the goodness of this inference is also contravened by natural principles governing good inference like Uniformity from Chapter 8.

[9] Kaplan (1989b, 540).

> Uniformity.
>
> The goodness of a deductive inference always depends solely on a relation between the contents involved in the inference, and some cognitive grasp of that relation.

The main forms of good inference of which we are aware don't differentiate between attitude states, and Uniformity is also well-motivated from a foundational understanding of what inference is, such as that given by even the skeletal picture of Chapter 2. But by this standard (1) cannot track a good inference, as it is transparently problematic to infer one's own existence without premises while engaging in a suppositional or imaginative task.

These quick remarks are not meant to be dispositive. The point is that treating (1) as modeling a good deductive inference would require substantial justification. But this brings us to point (B): claiming that LD fails to model good inference does not contravene anything that Kaplan says in "Demonstratives" or "Afterthoughts". In both documents, Kaplan is quite clear what he intends validity to track, and it is *not* framed in terms of deductive inference. For example, see the characterization of validity under (V) in the most recent quotation from Kaplan above. Or the following:

> Validity is truth-no-matter-what-the-circumstances-were-*in-which-the-sentence-was-used*. As I would put it, *validity* is universal truth in all *contexts* rather than universal truth in all possible worlds.
>
> (Kaplan, 1989a, 595)

Kaplan hardly mentions inference or reasoning in the course of his papers. Instead, he frames validity directly in terms of truth at all proper contexts. Inasmuch as we think of validity in these terms, I have no quarrel with LD. If this is what one means by "valid", then the sentence in (1) *should* be valid: it is true at all proper contexts, because at all of those a candidate speaking agent exists. It is, trivially, true at all circumstances where an referent of the first-person pronoun exists.

I should here again emphasize the genuinely ecumenical stance adopted in Chapter 1 that "logic" is a term of art, and one can use it to label many different kinds of properties or relations. And the property that Kaplan labels "valid" is, I think, a perfectly reasonable property for theoretical investigation. In fact, I would group Kaplan's form of validity within

a family of interrelated notions, each of which could be worthy of independent investigation. Kaplan focuses on sentences. But we could also focus on thoughts. We could investigate what it would be for a sentence to be one where the-thought-expressed-by-the-sentence-is-true-no-matter-what-the-circumstances-were-in-which-the-thought-was-held. This would yield as valid not only "I exist" but also (if the ability modal "can" and epistemic language "thinks about" can count as 'logical') valid sentences of the form "I can think about X" for any X. Necessarily, anyone who thinks the thought expressed by "I can think about the Revolutions of 1848, Carlos Santana, and the Law of Quadratic Reciprocity" is *ipso facto* able to think about the three things they believe they are able to think about. Something like this way of privileging sentences or thoughts has actually received some attention in the literature on self-knowledge. In a series of papers and a recent monograph, Alex Byrne has defended an idea (following up a suggestion of Evans (1982)) that we are able to know what we are thinking by following an 'epistemic rule' of the following form:

> BEL: if p, believe that you believe that p.[10]

As Byrne points out, this rule is *strongly self-verifying* in the following sense: if one tries to follow the rule—that is, if one follows the rule because one believes its triggering conditions obtain—then the belief that one acquires as a result of following the rule is *guaranteed* to be true.[11] Rules of the form "in any circumstance, form the belief that you can believe things about X" will also have this strongly self-verifying property for every X.

I think the sentences and rules just adumbrated form a very interesting class. They may even have a privileged epistemic status of some kind. For example, Byrne claims that by following BEL you can gain *knowledge* of your own mental states. Whether Byrne is right about any of this, we can take away two important points. The first is that relations like strong self-verification are perfectly reasonable to want to demarcate and investigate. But the second is that relations uncovered by strong self-verification do not appear to have any direct relationship to good deductive inference. Byrne's rule BEL, for example, is not a good way of *extracting information* from your starting acceptance state.[12] A

[10] Byrne (2005, 2011, 2018)

[11] Subject only to the minor caveat that counterexamples can occur if one 'changes one's mind' mid-rule-following.

[12] See Valaris (2011) for a criticism of Byrne's rule *as* a form of inference by appeal to Uni-

'logic' that says that "I think p" follows from "p" could be a very interesting one. But it would clearly not be a logic *for good deduction* in my sense.

Kaplanian validity and strong self-verification are just two members of the much larger family of interesting 'logics' one could investigate in this sphere. For example, turning back to sentences, we could investigate those which must be true whenever spoken, and so could include (with suitable accommodation of logical vocabulary) new 'validities' like "I am speaking now". Kaplan actually discusses a relation like this in "Afterthoughts"—sometimes calling it "utterance validity" and uses it to contrast with the notion of validity that is of interest to him.[13]

Again, this property of utterance validity doesn't seem useful in tracking relations of good inference. And, in fact, at most one member of the broader family of 'logics' I have been sketching could possibly track good deductive inference, since each member of this family tracks different relations.

So the lesson to take form point (B) is that Kaplan explicitly defined validity in terms that did not mention deduction or even reasoning. The terms in which he did define it have no obvious direct ties to deduction. Perhaps there *could* be such ties, but they would need to be argued for. And *there is no attempt in Kaplan's work* to provide any such argument.

Again, I regard none of this as coming into conflict with anything Kaplan claimed. That said, it *does* seem to come into conflict with the claims of philosophers inspired by Kaplan like Zalta or Russell. Russell says that Kaplan's logic shows that it *must* be a mistake to characterize logic as concerned with necessary truth-preservation, as I have done. But Russell does not supply arguments that would justify this sweeping claim—for example, by giving arguments that bridge Kaplan's particular treatment of validity with other plausible relations one could track using the term "valid". And surely a use of the term to track features of good deductive inference could count as one such relation.

Now, even if neither Kaplan nor those inspired by him have supplied con-

FORMITY.

[13] Kaplan (1989a, 584–5). Kaplan actually flags, very plausibly, that in fact "[t]here are different notions of utterance-validity corresponding to different assumptions and idealizations" (Kaplan, 1989a, 585,n.40). So even here we have a plurality of possible formal objects to study. He also notes that his definition of validity for true demonstratives starts to incorporate the resources that would naturally be used in defining utterance-validity (Kaplan, 1989a, 586,n.43). In connection with this last point, see the problems for Kaplan's logic for true demonstratives in Appendix C.

nections between his characterization of validity and good deduction, there is obviously an important further question about whether such connections could be supplied. This brings us to point (C): although *some* validities and consequences in LD do not seem to track good inference, many others do seem to 'correspond' to good inferences in *some* sense. Consider again (3) and (4).

(3) $\text{I} = James \models_{LD^-} N\textsc{located}(James, \textsc{here})$

I am James ∴ James is now located here.

(4) $\textsc{here} = Pittsburgh \models_{LD^-} N\textsc{located}(\text{I}, Pittsburgh)$

Pittsburgh is here ∴ I am located in Pittsburgh.

Imagine for the moment that you are afflicted with severe amnesia. You wake up not knowing who or where you are. You gain evidence that you are James, and so come to believe that you are. Can you now infer to a new belief, that you would naturally express by saying "James is now located here"? This seems like an *excellent* inference. Unlike the inference to one's own existence from *no* information, here the basing attitude that you are James seems informationally sufficient for drawing the conclusion in question. The same would go for an inferential analog to (4).

What of my complaint against (1)—that the goodness of the inference doesn't intuitively persist in imagination or counterfactual supposition? Doesn't that apply here too? Here things become somewhat complex. This much is true: it does not seem correct to *report* an intuitively good inference involving counterfactual attitudes using the indexicals in (3) and (4). If you suppose, counterfactually, that you are James, you should not inferentially transition to a supposition you would express by saying "I have further supposed that James is [or: were] now located here". That seems to report that you inferred that James is located where you are *actually*, while busy counterfactually supposing that you are James. And that would be a terrible inference. But, intuitively, this is a problem *with the words* you are using. That is, these words do not seem to describe the *same inference*—the same transition in informational content—that you would undergo in the imagined case one paragraph back while suffering from amnesia. Rather, it seems the indexicals are misfiring, and picking out a *new* piece of information that one could (erroneously) conclude.

Is it possible to express the otherwise intuitively good inference, transposed to supposition, in ordinary language? I think any temptation to try

would involve *other* indexicals. For example, after supposing you were James, you might express the conclusion of your inference under supposition by saying "I am now supposing that James is [or:were] *then there* (whenever and wherever that might be)". Kaplan familiarly also gave a semantics for demonstrative expressions like these. Without getting into the details here, their extensions are in part fixed by associated *demonstrations*. And here the demonstrations seem to 'reach into the content' of the initial supposition, as it were, to pick out the time and place of the first-personal supposing. Does this demonstrative trick work, and if so how? I won't take a stand on this here. More important is the idea that there *is* a good inference 'corresponding' (as I feel best to put it) to the consequences described in (3) and (4), and which survives even in imagination or counterfactual supposition. The problem is that it is not clear how best to *express* that inference using ordinary language, including the language of indexicals and demonstratives.

The foregoing remarks are obviously gestural. One thing I hope to accomplish with them is to motivate the idea that if we want to gain an understanding of what a specifically *inferential* logic for perspectival language would look like, we need to seek greater precision through a somewhat different approach than Kaplan took. In particular, we should start probing key questions which Kaplan did not: Is there a distinctively good form of deductive reasoning that governs thought that we seem to express with indexical language? If so, what is it? How does it relate to the logic given by LD? And why (as the latest examples suggest) might there be obstacles in characterizing any such form of good inference using language?

I think we can make headway in all these tasks by approaching the topics of perspectival thought in a somewhat more direct way than Kaplan did.[14] In particular, we can gain insight by investigating the perspectival aspects of indexicality not through the lens of language, but directly in thought itself by consulting the literature on what, following David Lewis, has come to be called "*de se* cognition".

[14] This is not to say that Kaplan ignored the self-standing topic of perspectival thought. See especially his remarks at Kaplan (1989b, 531) and Kaplan (1989a, 597). The key issue is that Kaplan always approaches questions about perspectival thought *through* language—which we will see is in danger of distorting the phenomenon.

10.2 Deduction and the *De Se*

Imagine a thoroughly inebriated celebrity seeing themself on screen being awarded a prize for acting. The celebrity is so confused they are unable to recognize themself so-displayed. And because of this, although they may be capable of forming a thought they would express with a demonstrative like "that person won the award," they have yet to form a thought they would most naturally express with "I won the award."

De se thought—thought about oneself in a distinctively first-personal way—is usually brought out by examples like these that aim to distinguish it from, or at least privilege it among, forms of *de re* cognition about objects. There is an important tradition, spearheaded by the work of Lewis (1979) and Perry (1979),[15] which takes *de se* cognition to pose a distinctive challenge to intuitive frameworks for modeling propositional attitudes like belief or desire. Philosophers embracing this idea are sometimes called "de se *exceptionalists*".[16] Exactly how exceptionalists frame the challenge posed by *de se* cognition, and what they take the upshot of the challenge to be, varies.

One of the things that I want to emphasize here is that the puzzles that motivate investigations into *de se* cognition need not be, and are in fact typically not, formulated as puzzles about *language*. Of course the puzzles will be framed *using* language. But the point is that these puzzles are not about the use of any particular expression like "I" or "now". To see this, let's consider some classic examples that kicked off contemporary investigations into perspectival thought.

Perry used the following case to motivate the idea that *de se* attitudes require us to posit a special class of attitude states, with distinctive ties to action.

> I once followed a trail of sugar on a supermarket floor, pushing my cart down the aisle on one side of a tall counter and back the aisle on the other, seeking the shopper with the torn sack to tell him he was making a mess. With each trip around the counter, the trail became thicker. But I seemed unable to catch up. Finally it dawned on me. I was the shopper I was trying to catch.
>
> (Perry, 1979, 1)

[15] With important antecedents in Castañeda (1966, 1967, 1968), and Frege (1918/1997).
[16] The terminology comes from Ninan (2016).

Perry claimed that when he learns that he himself is the shopper with the torn sack, he comes to believe a content he may well have already believed (*that Perry is the shopper with the torn sack*) but with a new kind of belief state that relates to that proposition—a first-personal one that has distinctive ties to agency. This is why, when he comes to believe the same content again, he behaves in a new way by stopping and fixing his torn sack. Note that Perry's case is formulated in such a way that if it raises any issues at all, those issues are not distinctively about the interpretation of *speech*.

Lewis, against Perry, took *de se* attitudes to require us to posit a new class of attitudinal *contents*, appealing to the following example.

> [Two gods] inhabit a certain possible world, and they know exactly which world it is. Therefore they know every proposition that is true at their world. Insofar as knowledge is a propositional attitude, they are omniscient. Still I can imagine them to suffer ignorance: neither one knows which of the two he is. They are not exactly alike. One lives on top of the tallest mountain and throws down manna; the other lives on top of the coldest mountain and throws down thunderbolts. Neither one knows whether he lives on the tallest mountain or on the coldest mountain; nor whether he throws manna or thunderbolts.
>
> (Lewis, 1979, 20–1)

Lewis's idea was that since the gods know all the 'worldly' contents there are to know, but they lack some further knowledge, that ignorance must be ignorance of a *new kind of object* of the attitudes. In particular, Lewis suggested we should enrich the objects of attitude states by accommodating some that can vary in their truth (or correctness) from time to time and person to person. On this view, when Perry learns that *he* is the shopper with the torn sack, he actually comes to believe a new proposition that he didn't believe before—one that is true 'of him,' but may be false of anyone else in the possible world Perry occupies. Again, note that whatever Lewis's case shows, it does not appear to be a lesson about *language*.

There are other approaches besides Perry and Lewis's—most notably the unusual proposal of Frege (1918/1997), on which each person's first-personal thoughts are inherently private and unshareable.

Not all philosophers have been moved by examples like those Perry and Lewis gave. Many philosophers, who we can call "de se *skeptics*", feel that any

challenges posed by *de se* cognition for a proper understanding of propositional attitudes are *equally* posed by other interesting forms of cognition with no distinctive ties to thought about the self.[17] For example, perhaps puzzles about *de se* attitudes are resolved as soon as we properly understand analogous puzzles for *de re* attitudes. Or perhaps a resolution of ordinary Frege puzzles, in which an agent can think about one object in more than one way, will also resolve any issues for understanding *de se* cognition as well.

This puts a logician of my stripe with an interest in indexicality in a tough place. To make any progress in understanding inference that corresponds to the use of indexicals, it is clear we must have a firm understanding of what *de se* cognition involves. Inference and reasoning, as I've been at pains to stress, are inherently mental processes or events. Language is of interest to us only insofar as it provides a clear window into the relevant mental activities. And, critically, the issues raised in the literature on the *de se* are framed in a way that is largely *independent* of theses about language. The questions about the *de se* are ones that can be raised, and apparently answered, without immediately taking a stance on how language works. So if we want to understand how perspectival *reasoning* works, we must settle questions about perspectival mental contents and mental states first. And that cannot be done without staking out some quite controversial non-linguistic commitments.

Regrettably, I do not have space to enter into the details of the debates between exceptionalists and skeptics and here. This means that in order to make any progress I will have to plump for a side in the ongoing debates between exceptionalists and skeptics, largely without justification. Even so it will be helpful, I hope, to see how a given stance on the questions raised by *de se* attitudes can have a direct bearing on logical matters.

My own views on the *de se*—defended in Shaw (2020)—follow a recent trend in the literature of what we can call 'second-wave' exceptionalism.[18] On the second-wave exceptionalist view, two things hold. First, while *de se* skeptics have largely been right to find the *original* arguments for *de se* exceptionalism wanting, there are nonetheless suitable modifications of those arguments that suffice to establish exceptionalism. Second, it is not clear that the modified, successful arguments that establish exceptionalism *privilege* Lewis's framework,

[17] See, e.g., Boer & Lycan (1980), Millikan (1990), Spencer (2007), Cappelen & Dever (2013), Devitt (2013), Douven (2013), Magidor (2015).

[18] Spearheaded by work like Ninan (2016) and Torre (2018).

or Perry's, or Frege's. Rather, the lessons for attitudes that arise purely from consideration of *de se* attitudes are more abstract.

On one way of casting the importance of *de se* cognition, the key lesson it reveals concerns obstacles to the utility of *shareability* and *transfer* of cognitive states with respect to the dimensions of time and agenthood. That is to say, a true or correct *de se* attitude would not necessarily have its truth or correctness preserved were the 'same' attitude state taken up by another agent, or the same agent at another time.[19] Each of Lewis, Perry, and Frege's accounts respect this constraint, though in different ways. For Lewis, shareability is violated because the contents of a *de se* attitude have accuracy conditions that vary world-by-world or time-by-time. For Perry, shareability is violated because attitude states determine correctness conditions in such a way that two agents, or one agent at different times, could be in the same belief-state with one believing truly and the other believing falsely. After all, the same attitude states could bear different contents with a shift in agent or time. And for Frege shareability is violated quite directly in that the content of a *de se* attitude is taken to be in-principle unshareable.

We can cut across these three different approaches by saying that what the puzzles raised by *de se* attitudes show is that characteristically first-personal attitudes of acceptance exhibit a *time and agent correctness-relativity*: if an agent at a time holds a *de se* attitude of this kind correctly, it does not follow that another agent in their world, or the same agent at a different time in their world, could hold the same attitude correctly. Attitudes that are not distinctively about oneself (or the time) do not exhibit this relativity: instead they exhibit only a *world correctness-relativity*: if an agent at a time holds an attitude correctly, it does not follow that an agent in another world could hold the same attitude correctly.[20] Note: the correctness-relativity of *de se* attitudes is not *necessarily* a relativity in content. For example, Perry's view is compatible with all contents of *de se* attitudes varying neither in correctness with respect to time nor agent. Instead, the correctness-relativity is located for Perry in the way an attitude state determines correctness by picking up different contents in different settings.

As I say, I cannot defend this claim about correctness-relativity in any detail

[19] Where sameness of attitude state is hard to describe precisely, but would hold of (say) physical duplicates.

[20] Note that as mentioned in Chapter 2, acceptance states other than belief, like supposition and imagination, are subject to a standard of correctness. The point I make here is not merely about belief, but all acceptance states.

here. The most I can do is to investigate the conditional question: if this *were* the moral of investigating *de se* attitudes, what would it teach us about good reasoning, and especially good deductive inference, in the context of holding such attitudes?

The first thing to note is that the time- and agent-correctness-relativity (in addition to ordinary world-correctness-relativity) of acceptance states means that such acceptance states determine a body of information given by agent/time/metaphysically-possible-world triples: those triples such that the attitude state would be held correctly relative to them.[21] Following custom, we can call each agent/time/world triple of this kind such that the agent exists at the time in the world a *centered world*.[22] And call the set of such centered possibilities determined by an attitude state the *centered information* associated with an attitude state. The centered information of an attitude state need not be its content. (For example, the centered information of an attitude state would generally *not* be its content on Perry's view.) Even so, it is a body of information we can helpfully associate with the attitude to mark its distinctive perspective-correctness-relative character.

It might be helpful to review the relationship between the centered information associated with an attitude state and the contents of those states on various approaches to *de se* cognition. For example, the centered information associated with the belief state held by someone we would report as 'thinking of themselves that they are James' would consist in the agent-time-world triples

$$\{\langle \text{James}, t, w\rangle \mid \text{James exists at } t \text{ in } w\}.$$

This is true on each of the Perrian, Lewisian, and a broadly Fregean approaches to *de se* cognition. In effect, this is what makes each of their frameworks adequate as responses to the problems of shareability and transfer raised by the *de se*.

But each of these views will have a different conception of the relationship between this centered information and the propositional content of the attitude states with which it is associated. On Perry's view, the content of a belief

[21] Lewis ultimately relativized content to *properties*, which I think can actually do better justice to the phenomenon of perspectival thought in certain tricky cases. But I will work with the multiply parameterized relativization here since it enables us to draw illustrative parallels to Kaplan's semantics for context-sensitive terms.

[22] The reason for focusing exclusively on triples where the agent exists at the time in the world is that these are the only metaphysically possible perspectives.

state held by someone we would report as 'thinking of themselves that they are James' would never correspond to its associated centered information. For me, the content would be what is expressed by the sentence "James is James"—a necessary truth. For any other agent N it would be the content expressed by the sentence "N is James"—a necessary falsehood. Note that these are kinds of content that could be thought by many others, without the help of characteristically first-personal attitude states. By contrast, for Lewis, the centered information associated with a mental state essentially gives its content. Lewis took the contents of *de se* attitudes to be properties—whose possession is also time-, agent-, and world-relative—and the correctness of the attitude to involve possession of the property. Accordingly, Lewisian *de se* contents can be, and often are, modeled by sets of agent/time/world triples where the agent exists in the world at the relevant time (what are sometimes called *centered propositions*). Finally, for a broadly Fregean treatment of the *de se* we could have either a Perrian or Lewisian elaboration, depending on how we treat the truth-conditions of the first-personal thought that one is James. Perhaps the content will have something like Perry's rigid truth-conditions, or perhaps they will have something more like Lewis's relativized structure. The integral feature of Frege's view is that these contents are not shareable. So *if* the content of the thought that one is James, when thought by me, has a metaphysically necessary profile of truth-conditions (as it would on Perry's view), it would still not be identical to the thought expressed by "James is James". After all that content *is* thinkable by many different agents and Frege's contents are not. The content I think must accordingly be a special way of thinking a metaphysical necessity, proprietary to me. Similarly, if the Fregean thought has a Lewisian relativistic content, it would still not (as Lewis held) be a proposition that could ever be the content of someone else's attitude state.[23]

The second thing to note is that *good deductive inference seems to require the preservation of centered information*. That is: it requires that every centered world at which basing accepting states would be correct is also a centered world at which a concluding acceptance state would be correct. This is even so if centered information comes apart from attitudinal content. As noted in §10.1, the

[23] Frege only held the unshareability thesis regarding 'agentially' centered thought, and not temporally centered thought. I think the arguments for *de se* exceptionalism require parallel treatments of agenthood and time (again, see Shaw (2020)). It is a little tricky to say what Frege would make of imagining being someone else. Is *this* also an unshareable attitude? I won't speculate further on the matter here.

inference from the belief that I am James to that of James being 'here' is intuitively a good one. If Perry is right, the content of my thoughts in performing this inference would be from the content expressed by "that James is James" to the content expressed by "that James is in Pittsburgh." It is not clear how the goodness of the inference could be secured by these contents. After all, another agent making an inference involving these contents would obviously be making a poor deductive inference. But even on Perry's view, the inference involves preserving the centered information associated with my belief states. The belief that I am James, for Perry, would be associated with information subsuming centered worlds whose agent is James. The belief that James is here is associated with those centered worlds where James is suitably proximate to the location of the agent of the centered world at its associated time. The first set of triples is a subset of the second. So the transition between the relevant belief states preserves associated centered information on Perry's view. And, as noted, it preserves centered information for Lewis and a Fregean as well: the associated centered information of the relevant states is the same for each of those theorists.

Some of this should sound *very* familiar. The explanation of the goodness of the inference sounds *almost exactly* like the explanation of why there is an entailment from "I am James" to "James is here" in LD. The same will hold for a whole host of intuitively good inferences involving *de se* cognition. And it is easy to see why. The way that an agent-time-world triple are metaphysically and representationally bound together in the relativized information associated with a *de se* attitude parallels the way that an agent-time-place-world quadruple are conventionally and metaphysically bound together in the function of Kaplanian character (conventionally, via the linguistic rules associated with indexicals; metaphysically, by the structure on possible circumstances at which speech acts can be produced).

It must be emphasized, though, that these are *mere parallels*. Despite the shared unity, Kaplan's system and any system describing conditions on good inference would be *different things*. One concerns a conventional, linguistic setting. Another concerns features of how mental states represent. This is why I've been speaking about intuitively good inferences 'corresponding to' entailments in LD. Kaplan's LD-validities and -consequences track certain properties of sentences through their relationships to the contexts in which they could be uttered. We could *redefine* validity and important attendant no-

tions in Kaplan's system to yield a logic concerning deductive inference. But redefinition would be necessary. For example, we would have to reinterpret the role of indexical expressions in the language: they would no longer model linguistic devices securing reference as a function of linguistic context; rather, they would 'mark the place' for aspects of an agent-at-a-time (or relations to some such agent-at-a-time) as components of centered information characterized by the remainder of the sentence in which the indexical is embedded. We would have to reinterpret characters from a linguistic convention governing expressions, to some feature of *de se* cognition—perhaps its content, if we are Lewisians, but perhaps something else like a guise under which a content is entertained. Then we could redefine validity in terms of 'necessary' correctness preservation across the centered information now corresponding to the reinterpreted sentences of the system (where the necessity in question quantifies over the triples relative to which centered information can be evaluated).

I've stressed that my claims in §10.1 against the possibility of using Kaplan's system *as* a system of inference needn't come into conflict with any statement he made. But I think recognizing the parallels between the system we are imagining and Kaplan's should give us pause. Again, I don't take issue with what one calls "logic". Many systems with quite divergent aims can be categorized under that heading. But a critical question is whether Kaplan's system *gains credibility* as a 'traditional' and important form of logic in part from the fact that it happens to track some important good deductive inferences.

Consider the following question: could Kaplan have defined proper contexts in different ways, and developed a logic which tracked truth at all proper contexts in those other senses? Kaplanian proper contexts are those in which a speech act *could* occur. But as noted in §10.1, we could instead focus on those in which a speech act *is* occurring (though where the utterance happening need not be one relative to which a sentence is evaluated). Why not explore this logic instead? Kaplan gives reasons for focusing on a notion of truth for a sentence-type/context pair, rather than for an utterance, construed as speech act: "Utterances take time, and utterances of distinct sentences cannot be simultaneous (i.e., in the same context). But in order to develop a logic of demonstratives we must be able to evaluate several premises and a conclusion all in the same context."[24] This is a reason to evaluate sentences-relative-to-contexts rather than utterances. But it is not a yet reason to consider the proper contexts, relative to

[24] Kaplan (1989b, 522)

which sentence truth is assessed, in the way Kaplan does rather than in the way I just suggested. Kaplan notes that focusing on contexts where a speech act is occurring will change what counts as valid: "there are sentences which express a truth in certain contexts, but not if uttered. For example, "I say nothing."" But as an explanation of what would be problematic with tracking the different class of 'validities' Kaplan only has the following, vague remark: "Logic and semantics are concerned not with the vagaries of actions, but with the verities of meanings."[25] What is this 'verity of meaning'? In what sense is "I am speaking" not a *privileged* verity of this kind, but "I exist" *is*?

Note that if we focus on a logic for perspectival deductive inference, rather than the linguistic setting, the formal choices are *forced* on us in a unique way by the choice of subject matter. For example, it is simply irrelevant to the investigation whether a speech act could be produced, or is being produced, or whether a thought is being held, and so on. Both the parameters relative to which correctness of a sentence is evaluated and their relationships are foisted on them by an independent standard: the standard of information preservation for *de se* cognition.

I do not intend for the point of these remarks to be fully transparent yet. We will return to think about them in more detail in §§10.5–10.6. Still, it is worth bearing these issues in mind as we turn to an important remaining question. I've said LD has the right kind of structure to model good inference for *de se* cognition *in many cases*. But for all cases? In considering this question, I will focus on Lewis's framework for *de se* cognition, but the lessons will generalize.

Lewis initially suggested that properties were not only necessary to model mental contents, but also seemingly took them to be sufficient. But, following a familiar and important line of criticism developed by Markie (1984) and Nolan (2006), I take there to be an important obstacle to that proposal. As Nolan notes, while Lewis's framework seems sufficient to model the contents of beliefs, it is seems insufficient to model the contents of other attitude states like desire. For example, it seems possible to desire that one never have existed while otherwise things would be similar to how they actually are. To model the content of this desire as a property would seem to require the property to belong to an object at a world something like ours. But wouldn't this possession of the property require the object to exist at that world, thereby getting

[25] Both quotations at (Kaplan, 1989a, 584–5).

the desire for non-existence incorrect? (This is especially a worry for Lewis, who conceived of properties as sets of world-bound possibilia.) Alternatively, what centered worlds could we use to model the content of the desire? The existence of the 'centers' (agents-at-a-time) of centered worlds is hard-wired into their structure.

So neither of these related approaches seems to get the content of the desire for non-existence correct. We needn't actually focus on desire to make the point (as Markie's version of the objection goes). Consider ways in which we can suppose or imagine Lewis's case of the two gods. Letting w be a complete 'centerless' description of what the world of the two gods is like, it seems one could imagine this in at least four ways, corresponding to the following instructions.

(i) Suppose things are as in w, and that you are the god on the tallest mountain.

(ii) Suppose things are as in w, and that you are the god on the coldest mountain.

(iii) Suppose things are as in w, and that you are one of the two gods (but don't suppose you are a particular one yet).

(iv) Suppose things are as in w, and that you are not there.

Lewis suggested that we model what he called '*de dicto*' content—content that intuitively is not distinctively about the self—as just more self-locating belief whose associated information didn't happen to distinguish among world-mates. As he put it: "Belief *de dicto* is self-locating belief with respect to logical space."[26] If we use agent-time-world triples to model attitudinal content, the content of a *de dicto* attitude would be a set of triples such that whenever $\langle a, t, w\rangle$ is in the set, so is $\langle a', t', w\rangle$ for any a' existing at t' in w. But *both* of the contents attributed in (iii) and (iv) above appear not to distinguish between world-mates. And each concerns the same 'centerless' world. There accordingly seems to be a worry that Lewisian *de dicto* contents conflate *neutrality* on who occupies the role of a distinguished perspective within a world according to some information with *rejection* of the existence of such a privileged perspective according to the information.

[26] Lewis (1979, 522).

I am less concerned here with the 'nature' of *de se* contents here (e.g., whether properties suffice for those purposes) than whether we have *enough* pieces of information to model the information associated with attitude states. Given this aim, a simple way around the problem is to enrich the space of entities relative to which we can evaluate the correctness of the information associated with an attitude state. In addition to centered worlds—which, recall, are agent-time-world triples where the agent exists at the time in the world—we can consider *agent-free centered worlds*: triples consisting of a null element †, a time, and a metaphysically possible word. These entities can be used to represent the non-existence of a privileged perspective-determining agent at the given time in the given world.[27,28] The information associated with an attitude state more broadly can then be represented as a subset of the set of all centered *and* agent-free centered worlds. I will call such a set of worlds a body of *center-neutral information*. There may be other ways of getting the results we want.[29] The important point is that if we focus only on *belief*, there will be a temptation to think of all alternatives as centered since it is grossly irrational (if even possible) to believe one does not exist in a distinctively first-personal way. But that shouldn't preclude the possibility of *representing* one's own non-existence within a world, even in the presence of a perspective on it.[30]

The above concerns have historically been raised for Lewis's treatment of

[27] Cf. Lewis's treatment of 'possibilities' for tuples of agents using null elements in Lewis (1980, 28).

[28] I've retained a time parameter since I take it to be possible to imagine a particular time's being 'now' within a given world, while imagining that 'oneself' does not exist (and indeed never exists) within that world. If that is wrong, further adjustments would be needed. I am also presuming that if one imagines oneself existing in a world one can be *neutral* on which time it is within the world at which one exists, but not that there is no such time. Accordingly, there is no added need for a null time parameter to handle such cases. Again, if this is wrong, further adjustments would be called for. I won't explore how these adjustments would go here.

[29] See Turner (2010) and Feit (2010) for slight adjustments or reinterpretations of a broadly Lewisian framework to account for the representation of non-existence.

[30] Can a single piece of information associated with an attitude state allow for both centered and agent-free centered worlds among its correctness-conditions (as I have opted for here), or must it either subsume exclusively centered worlds or subsume exclusively agent-free centered worlds? I think the former option is bolstered by the idea that if I first ask you suppose the world is some way, I can then further ask you to suppose that you are within it, or I can ask you to suppose further that you are not. Either way involves a *specification* of the initial information, which requires that the initial information contain both centered and agent-free centered worlds. Independently of this, 'mixed' bodies of information get what I take to be the right results for denying existence entailments within a logic modeling inferential connections for *de se* cognition, discussed below.

centered cognition, but of course the lesson here generalizes to other treatments of the *de se*. We want to be able to capture the sense in which the information associated with an attitude state can encapsulate the non-existence of the perspective of an existent, whether or not we take up Lewis's particular treatment of what having that perspective involves. And I should stress again that the 'mixed' center-neutral bodies of information that I think we should associate with attitude states need not be the *content* of those states, as a broadly Lewisian approach would require.

This broadening of the type of information associated with *de se* cognitive states suggests an adjustment to any logic that aims to model good inference. Good inference preserves correctness across the correctness-evaluable elements of the information associated with an attitude state, which now include not only centered worlds which are already helpfully captured by proper contexts within $LD^{(-)}$, but agent-free centered worlds as well. To accommodate the latter, we simply need to broaden the formal counterpart to proper contexts. This would simply be done by changing clauses (I) and (VIII) of the definition of an LD^{-} structure from page 295.

(I′) $\mathcal{C}$ is a nonempty set of contexts $c = \langle c_A, c_T, c_P, c_W \rangle$ with:

(i′) $c_A \in \mathcal{U} \cup \{\dagger\}$ (the agent of c),

(VIII′) If $c \in C$, then if $c_A = \dagger$, $\langle c_A, c_P \rangle \notin \mathcal{I}(c_W, c_T, Located)$, and if $c_A \neq \dagger$, $\langle c_A, c_P \rangle \in \mathcal{I}(c_W, c_T, Located)$;

Kaplan's original (VIII), in conjunction with (IX), secured all contexts as proper ones in Kaplan's sense. (VIII) ensured contexts were such that the agent of the context is located at the place of the context at the time of the context. With (IX) this also ensured the agent existed at the relevant time and world. By severing the locational claim for the null element † in (VIII′), we prevent its existence from being secured in this way.[31]

With these modifications, we can keep the remainder of Kaplan's definitions—including those of validity and consequence—in place, and call the resulting logic LD^{*}. *This is the logic we obtain for modeling conditions on*

[31] There are obviously further questions about handling the 'place' of a context with the shift of center-neutral bodies of information, which I haven't considered here and which might lead to a different conception of the information associated with an attitude state and different associated modifications of LD. Can one imagine a place being 'here' in a world even if one doesn't exist there? I suspect so. But I won't investigate the issue here.

good deductive inference for distinctively perspectival thought in an exceptionalist framework.

What changes in the transition to LD^*? Notably, problematic existence entailments like those in (1) and (2) have disappeared.

(1) $\not\models_{LD^*}$ EXIST(I)

I exist.

(2) $\not\models_{LD^*}$ NLOCATED(I, HERE)

I am presently located here.

This is simply because it is no longer guaranteed that an agent exist at a 'context'[32] in LD^*. But we keep the virtuous inferential connections that depend on the way that features of a perspective are bound together.

(3) I $= James \models_{LD^*}$ NLOCATED($James$, HERE)

I am James ∴ James is now located here.

(4) EXIST(I)∧HERE $= Pittsburgh \models_{LD^*}$ NLOCATED(I, $Pittsburgh$)

I exist and Pittsburgh is here ∴ I am now located in Pittsburgh.

The consequence in (3) now holds because any 'context' at which the premise is true is one where "$James$" and "I" both refer, and corefer. Any such context is one where "EXIST(I)" is true, and as a result clause (VIII′) will impose a structure on this context ensuring "NLOCATED($James$, HERE)" is true at it. (4) holds because "EXIST(I) ∧ HERE $= Pittsburgh$" can only be true if the agent of the context exists and the place of the context is Pittsburgh. Again, clause (VIII′) will ensure of any such context "NLOCATED(I, $Pittsburgh$)" is true.

In short, what an inferentially-based logic guarantees is that *if* information is given from an existent's perspective, *then* that existent is integrated with elements of the perspective (time and place) in a certain coordinated way. What logic does not guarantee is that information is given from the perspective of an existent to begin with.

These are good first steps to developing a logic that models good deductive inference in the context of perspectival thought. There are many matters

[32] Bear in mind these are now modeling something quite different than for Kaplan.

to consider in developing it further, but I do not want to pursue the relevant nuances here. The adjustments I've made so far suffice to make a number of conceptual points that matter to me most here.

First, it is worth stressing that the redefinition of a context in LD^* could not have been motivated on Kaplan's characterization of validity. If validity is, as he puts it, "truth-no-matter-what-the-circumstances-were-in-which-the-sentence-was-used", then it would be a *mistake* to characterize contexts as possibly involving non-existent agent parameters. Kaplan developed the correct logic for *his* characterization of validity. But that also means that it was not incidental that Kaplan's logic does not model good deductive perspectival inference—his conception *precluded* modeling it. Truth-no-matter-what-the-circumstances-were-in-which-the-sentence-was-used is not only conceptually, but extensionally distinct from 'perspectival-information-that-can-be-inferred-without-premises.'

Now that we have the rudiments of a perspectival logic for deductive inference we can finally return to earlier worries about the failure of necessitation. Do these failures (which persist in LD^*) show, as Russell puts it, that a "definition of logical consequence [on which] a sentence A is a logical consequence of a set of premises Γ if and only if *it is impossible for all the members of* Γ *to be true and* A *false*…is a mistake"? Here it should be clear that a key lesson of Chapter 8 reapplies. While there remain validities in LD^* for which necessitation fails, it is transparent that this is merely because the language-internal necessity operator is not tracking the correctness of the information associated with the *de se* attitudes which the sentences of the logic has been recruited to model. This information admits of a characterization in terms of necessity (correctness at all center-neutral worlds) which is precisely the notion which matters to the goodness of deductive inference. As before, we could develop a language-internal necessity operator that, merely as a formal matter, captures this notion. This operator would have to shift and bind parameters of LD^*-contexts in a way that Kaplan thought never occurred for the linguistic contexts of LD.[33] But whatever the prospects for introducing such an operator in a spoken language, the point will remain: there *is* a generalization of metaphysical necessity that undergirds good deductive inference, and a logic investigating such inference

[33] See Kaplan's remarks on 'monsters' in Kaplan (1989b).These have given rise to some controversy—e.g. in Schlenker (2003), Santorio (2012), and Rabern (2013)—that need not concern us here.

should be set up precisely to track it.[34]

It might be objected that this form of necessity does not govern the *content* of the attitude—at least on certain views like Perry's. But even if this were true, it would not matter. For on a view like Perry's we should simply recognize that there are two broadly truth-conditional bodies of information associated with an attitude state, each of which involves a range of metaphysical possibilities: the center-neutral information associated with the attitude state, and the attitude state's propositional content. It is plain at that point that there are two notions of necessity to consider when asking whether logical truths are necessary—*even if* we require the necessity in question to be a metaphysical modality. After all *both* bodies of information subsume information about metaphysical modality. What is more, there is no special pressure to privilege one of these bodies of information because it is the content of the attitude state provided it is demonstrable that such content is not primarily relevant to our aim of tracking good deduction. And this *would* be demonstrable in this case.

This means that far from presenting a problem for the idea that logic concerns itself with a form of necessity, investigation into the role of perspectivally governed information actually provides additional support for it. This doesn't mean that we have nothing to learn for the foundations of logic from such an investigation. On the contrary, we can now appreciate one definite lesson of this kind, as well as space for a further possible lesson.

The definite lesson is that developing a logic for perspectival information requires acknowledgment that the notion of necessity with which good inference, and so logic, is concerned is not *mere* metaphysical necessity—truth at all metaphysically possible worlds.[35] Rather, it is a necessity governing information about metaphysical possibilities that *may* integrate a perspectival element

[34] This may seem to involve a mere formal trick—introducing a new kind of necessity operator—that could have easily been executed *before* the transition from the logic LD to the logic LD^* for deductive inference. But there is a conceptual point that undergirds the formal change. While we can of course quantify over contexts in the Kaplanian sense and develop a 'notion of necessity' which validity in his sense tracks, the resulting notion is *not* a plausible generalization of necessity in its role in undergirding good inference in the non-perspectival case. This is because contexts in the Kaplanian sense, in spite of being formal 'refinements' of metaphysical possibilities, are not conceptually tied to correctness-determination of cognitive information. They are mere representations of speech act settings. This is not true of the tuples playing the role of 'contexts' for LD^*, which are introduced precisely as parameters with respect to which the information associated with a cognitive state can be evaluated for correctness.

[35] Well, at least this is a definite lesson *given* a broadly exceptionalist treatment of *de se* cognition I've taken on as a background assumption of this section.

that introduces grain to metaphysical modality.

The second, merely possible, lesson concerns the relationship between the bodies of information logic studies and assertoric and mental contents. If Kaplan is correct about what the contents of speech acts are, then these contents do not have a relativity to perspectival parameters like agenthood built into them, and so the information that (even a deductive inferential) logic investigates is not *generally* that given by the speech act contents of assertion. (This is one interesting conclusion distinctively about logic that Kaplan emphasized, and that nothing I have said so far contravenes.) Kaplan did not merely take his claims about assertoric content for granted, but argued for them directly. And while not all philosophers were persuaded by his arguments, Kaplan's views have a good right to stand as the orthodox position.[36] What is even more surprising is that if theorists like Perry are right, this is not merely a result of a mismatch between assertoric and mental content. Rather, mental contents themselves may not be perspectival. This last view is much more controversial. But even assuming *de se* exceptionalism is granted, *de se* skeptics have helped to reveal that we are only just coming to correctly appreciate its grounds. Accordingly, a view like Perry's can hardly be ruled out at this juncture.

Both types of lesson are intriguing. Each requires us to rethink some aspect the representation involved in inference and the ways in which we could capture it through linguistic modeling. But it is important to stress that neither lesson casts any doubt on the idea that there is *some* body of broadly metaphysically modalized information whose necessity regulates good inference, and that a reasonable conception of logic can be concerned with trying to track it.

Let me recap some of the key points of the last two sections before moving on. Kaplan developed the logic LD on the basis of investigations into the compositional semantics of perspectival context-sensitive terms and the assertoric content expressed with their help. This logic is excellent at capturing validity as Kaplan conceived of it: as a guarantee of sentence truth provided by standing linguistic meaning *given* a metaphysically possible linguistic context in which the sentence could in-principle be produced. But this notion of validity does not, and cannot, model good inference in the presence of perspectival representation. Doing this requires taking a stance on puzzles about mental

[36] See Lewis (1980) for some doubts about the intuitions Kaplan appealed to. See also the growing literature (e.g. in Ninan (2010a), Maier (2016), Caie & Ninan (forthcoming)), building on a concern of Stalnaker (1981), for trying to understand how communication of perspectival information can be accounted for given *de se* exceptionalism.

representation of this kind. And important exceptionalist accounts of such representation reveal bodies of information associated with perspectival mental states governing good deduction that Kaplan's framework does not, and cannot, model consistently with the interpretation Kaplan gave to logical notions. Still, a set of languages and a logic like that Kaplan developed can be reinterpreted and modified in slight ways to model the relevant features of good deductive inference. The resulting logic invalidates the suspicious existence entailments of LD, as we should have expected, but it also replicates the logical connections Kaplan forged between aspects of a perspective like an agent, a time, a location, and a world. It also makes clear that the persistent failure of necessitation (for the style of necessity operator Kaplan defines) may still help to reveal interesting features of the information logic concerns itself with, such as that such information may not stand as the assertoric content of declarative sentences of a natural language like English. But the failure of necessitation does not reveal that a logic for good deductive inference does not concern itself with relations of necessary truth-preservation. On the contrary, the proper formulation of such a logic for deduction reveals exactly the opposite.

10.3 Good Inference and the Passage of Time

Before moving on I want to briefly discuss a topic that arises naturally in the current context: how is deductive reasoning about *time* possible if inference requires the passage of time (both intuitively, and as my analysis of inference from Chapter 5 supports)?

With this question in mind, it is intriguing to reconsider Kaplan's remarks about why he does not relativize truth to an utterance: "Utterances take time, and utterances of distinct sentences cannot be simultaneous (i.e., in the same context). But in order to develop a logic of demonstratives we must be able to evaluate several premises and a conclusion all in the same context."[37] But why *must* we consider multiple claims—'premises' and 'conclusions'—in the same context? Obviously we would have to consider multiple claims *if* logic was concerned with something like a process of deduction, since deduction requires separate consideration of premises and conclusions. But Kaplan does not characterize logic in these terms. And the terms in which he *did* characterize validity are not ones that make it at all clear why logic would ever be

[37] Kaplan (1989b, 522)

interested in a *relation* between sentences (or their contents, etc.), as his quotation suggests is of some importance. Finally, even if Kaplan had cared about deduction, *actual* deduction takes time and, necessarily, cannot occur in a single context in Kaplan's sense. What would we learn about deduction using idealizations that seem to model *no metaphysically possible* instance of it?

I think these questions reveal that Kaplan's stated motivation for not relativizing truth to an utterance betray some ambivalence about the nature of logic. Kaplan seems to state that logic should care about a relation, though he never explains why, nor does he ever even define the relational logical notion in question. The clearest motivation would be a concern with deduction, which (a) Kaplan never explicitly discusses, (b) is not obviously compatible with Kaplan's explicit characterization of logical validity, and (c) does not obviously support his use of the relational character of logic to focus on sentence truth at a single context anyway.

But let me set Kaplan aside for now. What should a theorist like me, who would use a logic like LD^* to model good deductive inference, say about the importance of temporal passage for a logic? It may seem to call for some sort of modification to the existing framework, or at least some acknowledgement of idealization. Perhaps surprisingly, however, no such modifications or concessions are called for. To see this, it helps to break inferences into two categories: inferences under hypothetical attitudes like counterfactual supposition and imagination, and inferences under belief.

Inferences under hypothetical attitudes, like all inferences, require the passage of time. But a moment's reflection reveals that here the passage of time raises no troubles for LD^*. This is because even if time must pass as a given inference is performed, this does not require any passage of time 'within the states of affairs' represented by information in the inferential transition. Indeed good deductive inference in this context would seem to require that there is *no such passage* in what is represented.

For example, you could suppose you ('*de se*') are someone at a train station at *exactly* noon with a train to your left and a train to your right. It would seem to follow within this imagining that 'you' have a train to your left at exactly noon. The inference here using conjunction elimination may take time. But no time 'passes' concerning the subject matter of your inference, and it seems clear that it *must not* to count as a good logical inference. Even if your inference took two seconds (say), it would not be a good idea to deductively

conclude under supposition that you must have a train on your left at 12:02. Relatedly, there is value in being able to say what follows from the initial perspectival information in your supposition state with the perspective held perfectly fixed. All this shows that we *need* a logic that holds constant even the parameter of time when investigating good deductive inference for perspectival information. That is, we need a logic which is something like LD^* just as it stands.

In fact, something like this holds even of the second category of inferences which mediate between beliefs. This can feel surprising. If it is exactly noon, and I go on to infer under belief that it is exactly noon from the fact that it is exactly noon, I could have transitioned from a true to a false belief provided the inference took enough time. The key is to note that while this would clearly be problematic, it would not obviously be problematic for the deductive inference, qua *deductive inference*. Here is a place where the resources for understanding the normativity of logic for deduction from Chapter 3 come back into play. Recall that to say a deductive inference is a good one is to say *nothing* about whether it is rational to perform it, whether one should perform it, or whether one has reason to perform it. Those facts will depend at least in part on whether one ought to, or has good reason to, perform an act of reliable information extraction. We should condemn the inference just described that concerns the time being exactly noon, but not because it is a bad deductive inference. We should condemn it because it would foreseeably lead to a false belief because it is a good deductive inference performed over contents involving overly precise times.

Note it is the precision that matters. If I infer from the time's being roughly noon to the time's being roughly noon, this is again a good purely deductive inference involving perspectival information which need exhibit no irrationality (beyond, perhaps, the irrationality of wasting my time performing trivial inferences). What this means is that good inference involving perspectival information about 'the present' must be exercised with some rational caution. But the norms governing that rational caution are not those belonging to a theory of good deductive inference *qua* inference, and so not to logic (as I conceive of it) either.

Could there be other forms of inference that take into account the passage of time? If I know that my inference will take me exactly one second, can I deductively infer that it is 12:01AM from the fact that it is 12:00AM? I think

not. I don't doubt that one can rationally transition between beliefs about the time like this in *some* way as a manner of registering the passage of time. But I see no reason to group that kind of rational transition in with those I have been calling "deductive inferences".

Consideration of both belief and other attitudes thus reveals no problem for a conception of deduction that requires holding temporal parameters constant. This is so even if, as I have maintained, it is *essential* to deduction that it requires the passage of time. This passage simply has nothing to do with the aim which gives deductive inference its attendant standard of goodness. And pursuing the aim of deductive inference is still perfectly reasonable in spite of the passage of time, both for belief and especially for hypothetical attitudes. So there is no reason to tinker with a logic like LD^* on account of time.

10.4 Logics for Strong Lexical Ambiguities

In §10.2, I defended the claim that a logic for deduction involving perspectival thought was conceptually connected to a generalization of metaphysical necessity. What became of Kaplan's claims about the contingency of the assertoric contents associated with some logical truths in his framework for perspectival context-sensitivity? I claimed that a logic in my sense would, provided Kaplan was right about the nature of the assertoric contents, simply ignore them as irrelevant to the information pertinent to perspectival deduction.

This may have been well and good for the investigation of perspectival inference, which proceeded in abstraction from language. But that abstraction from language raises important lingering questions. If the assertoric contents of perspectival indexicals end up not being relevant to the information in perspectival mental states, then is there *no* logic for linguistically context-sensitive expressions at all? If so, why not? But if there is a logic for those context-sensitive expressions, what would it be? *Couldn't* it be LD after all? And if so, wouldn't many or all of Kaplan's claims about logic stand anyway?

I will begin to address these questions in §10.5. But to properly situate my ensuing remarks, it will be extremely helpful to take an extended detour to discuss the third of the phenomena I previewed that I would investigate in this chapter: lexical ambiguity. The main reason for this investigation is instrumental, as I will eventually claim that we should expect logical parallels between lexical ambiguity and context-sensitivity, and indeed that we find them. What

is more, the case of lexical ambiguity is complex and illuminating to consider in its own right.

So, given all this, let's ask: what would a logic in the presence of lexical ambiguity look like?

Roughly, an expression type is ambiguous if it has multiple interpretations that trace to distinct linguistic conventions. I will focus here on lexical ambiguity (as opposed to syntactic ambiguity), in which the variation traces to different conventions governing lexical entries that are homonymous or co-spelled. This is exemplified by the English "bank" which can refer to a financial institution or a river's edge. Unlike with ambiguous terms, shifts in extensions among context-sensitive terms like "I" stem from a single linguistic convention (in this case, that its denotation be the speaker of a context).[38]

There are arguably some ambiguous terms whose divergent uses nonetheless have close semantic connections. Words labeled 'polysemous,' which have several distinct but semantically related meanings, would be good candidates. Consider that the word "in" as used in "my keys are in the drawer" and "you are always in my thoughts" seem abstractly semantically connected—both are concerned with 'inclusion' in some sense. But the uses are different enough that it would not be surprising to see them translated differently into languages other than English.

I mention such cases to set them aside. The case that is of more immediate interest is that where ambiguous language seems semantically unrelated like that of "bank". Pairs of such homonyms or co-spelled words are each usable as a means of expressing mental states, but with semantic differences significant enough that is clear that they share no interesting common inferential properties (beyond those that they would share with any other member of their syntactic type).

When a lexical ambiguity is such that two uses of a term share no semantic connections relevant to inference in this way, I will call it a *strong ambiguity*. The question I want to focus on is: What should we do if we want to investigate the inferential logic of a language in which such strong ambiguities are present?

Consider (9), which appears to report an inference.

(9) Rashid is at the bank ∴ Rashid is at the bank.

[38] See Sennet (2016) for a discussion of various forms of ambiguity and their contrasts with related phenomena like context-sensitivity, polysemy, vagueness, underspecification, and sense-transfer.

Does it express a good inference? Well, obviously, one can reasonably infer that Rashid is at the water's edge from his being at the water's edge. And one can't infer that Rashid is at a financial institution from the fact that Rashid is at the water's edge. So we have at least one good inference, and at least one bad inference, each expressed by co-spelled sentences. What to do?

The first option is both the simplest and by far the most common: we could investigate the logic of the language when ambiguous terms are fully disambiguated. This can be done in several ways. We could simply translate the language into a different language that is free of ambiguity; or we could 'annotate' the existing language (e.g. with different subscripts to reflect different interpretations); or we could relativize consequence relations to information about disambiguation—for example by pairing each ambiguous expression with a function from numbers to its possible interpretations, representing which interpretation to give the expression on its nth occurrence as we read through from premises to conclusions; and so on.

But if we don't take one among this class of options, our logic will relate ambiguous sentences without either overt or tacit disambiguating information. What could this logic model? It cannot directly model necessary preservation of truth, or even actual preservation of truth, as a condition on good inference since ambiguous sentences don't rise to the level of expressing content which could be true or false.[39]

Then again, in some sense, *most* logics do not characterize good inference directly. On the picture advanced in Chapters 2 and 7, the characteristic feature of a logic is that it abstracts from certain linguistic properties to effectively demarcate *classes* of inferences that share some revealing similarities. Sometimes logicians even go so far as to attribute logical properties to objects that are not candidates for truth-evaluability precisely to highlight commonalities among members of a given class. For example, we can attribute logical properties to *logical schemas*, in virtue of patterns of truth among their instances.[40] Schemas themselves say nothing, and are neither true nor false. Their utility comes by way of classifying groups of sentences that are true or false into categories. Indeed, I pursued something like this strategy in Chapter 7 by permitting *uninterpreted* first-order sentences to be bearers of validity.

So couldn't ambiguous sentences play a role something like schemas do? It

[39] Cf. Lewis (1982, 438).

[40] Cf. Quine (1970/86, 50–1)

turns out that there *is* some sense in which they can. The key issue is that when singling out classes of inference in the ambiguous setting an abnormality may creep into the characterization of classes of inference that we must be mindful of. To see why, let's start to explore what the classes would look like.

Once we forbid a logic from accessing information about *actual* disambiguations, we can preserve contact with truth by considering truth on, or truth-preservation across, *possible* disambiguations by quantifying over them. When we do this we can vary a conception of consequence along several dimensions including:

(a) quantificational force (e.g., we can focus on truth on *all* disambiguations, or merely on *some*);

(b) quantificational scope (with most narrow scope over a single sentence, or widest scope over all sentences and any conditional used to express truth-preservation); and

(c) kinds of disambiguations considered (uniform, in which we always disambiguate the same ambiguous type the same way; or 'mixed,' in which allow a single type to be disambiguated multiple ways).

These choices give rise to a family of relations, loose specifications of which are below. I label these using $\forall/\exists$ to track quantificational force, N/W to track narrow or wide scope of quantification with respect to a conditional in the entailment relation, and U/M to track uniform or mixed disambiguations.

$\exists$NM: If every premise is true on some mixed disambiguation, so is the conclusion.

$\forall$NM: If every premise is true on all mixed disambiguations, so is the conclusion.

$\exists$NM+$\forall$NM: If every premise is true on some mixed disambiguation, so is the conclusion, and if every premise is true on all mixed disambiguations, so is the conclusion.

$\exists$NU: If every premise is true on some uniform disambiguation, so is the conclusion.

$\forall$NU: If every premise is true on all uniform disambiguations, so is the conclusion.

∀WM: For each mixed disambiguation: if the premises are true on that disambiguation, so is the conclusion.

∀WU: For each uniform disambiguation: if the premises are true on that disambiguation, so is the conclusion.

We could obtain many more logics by continuing to vary the relevant parameters.[41] Lewis (1982) highlighted that relevance logics provide ways of modeling some of these relations. For example, in the first-order setting ∃NM corresponds to the first-order fragment of the logic LP of Priest (1979a), and ∃NM+∀NM to the logic R-mingle of Dunn (1976a,b).

This, of course, still leaves open *why* one would investigate these logics, and what limitations they might have. Here I want to focus on ∀WU for two reasons. First, this relation will suffice to highlight a general lesson for how all logics for ambiguity can in principle specify problematically gerrymandered classes of inferences. Second, we can learn some specific lessons about how ∀WU 'undergenerates' with respect to non-ambiguous languages that will provide a helpful contrast for logics of context-sensitive expressions. Let me take these points in reverse order.

∀WU generates an especially well-behaved logic. In the first-order setting, assuming conditions that would otherwise lead to classical logic in the absence of ambiguity (see Chapter 7), the relation $\models_{\forall WU}$ will extensionally coincide with what would otherwise be classical entailment relations over first-order sentence types. This is because every uniform disambiguation of a sentence in a language with ambiguities corresponds to a first-order interpretation of that sentence in a non-ambiguous language. (After all, an ordinary first-order model would interpret all repeated instances of non-logical vocabulary uniformly.) This can lead one to the following question. Since $\models_{\forall WU}$ relates the very same sentences as classical consequence, aren't these relations giving us the *same* information about the goodness of various deductive inferences? (Cf. my claim from Chapter 9 that certain forms of Strawson entailment simply *are* classical entailment relations, rather than merely being coextensive with them.)

The answer is that $\models_{\forall WU}$ does not merely re-specify classical inferences. This is in part because, unlike the classical consequence relation, $\models_{\forall WU}$ is relat-

[41] For example, we could further vary quantificational force (e.g., by considering truth on a single disambiguation, or on most), scope (e.g., having scope be wide over premises, but narrow with respect to the conditional), and uniformity (e.g., considering disambiguations that are uniform within a sentence, but mixed *between* sentences).

ing potentially ambiguous types. As such, the relation $\models_{\forall WU}$ can in-principle undergenerate with respect to a classical consequence relation along at least two dimensions. The first arises because $\models_{\forall WU}$ requires disambiguations to be uniform. The second arises because it quantifies universally over disambiguations.

To understand the first kind of undergeneration, consider two interpreted languages: a first language free of ambiguity, and a second language with several ambiguous expressions. The former univocal language contains sentences (10i-ii) that translate directly to (10i′-ii′) in the latter. That is, the three distinct names a, b, and c in the first language are all co-spelled by the three-way-ambiguous A in the second language.

(10) (i) $a = b$ and $b = c$.
(ii) $a = c$.
(i′) $A = A$ and $A = A$.
(ii′) $A = A$.

With identity as a logical term, (10i) classically entails (10ii) witnessing the transitivity of identity. Now, again provided identity is treated as a logical term, (10i′) entails (10ii′) under $\models_{\forall WU}$ as well. But the problem is that the interpretation of (10i′–ii′) that is of interest—the one which manifests the logical property of transitivity—is not one on which ambiguous terms are uniformly disambiguated. Accordingly, the relation among *ambiguous* types given by $\models_{\forall WU}$ 'says' nothing about the sentences (10i′–ii′) given a mixed disambiguation. Note, for example, that while the premise given by (10i′) (as ambiguous type) entails (10ii′) under $\models_{\forall WU}$, so does the empty set. This is obviously not true of the translations of the (disambiguated) conclusion back into our first, ambiguity-free language.

A speaker of the ambiguous language can wonder the *very same* thing that a speaker of the univocal language wonders when the latter considers whether they can infer what (10ii) expresses from what (10i) does. The speaker of the ambiguous language would express what they thereby wonder by asking whether they can infer what is expressed by the (disambiguated) (10ii′) from the (disambiguated) (10i′). Since in the ambiguous language this good inference can only be expressed using a mixed disambiguation, and since $\models_{\forall WU}$ has foregone the resources to characterize mixed disambiguations, it has foregone the resources to describe this particular inference's goodness.

So even though $\models_{\forall WU}$ relates the same first-order sentence types that a classical consequence relation would, because the former tracks a relation among ambiguous types it still undergenerates relative to the consequence relations for univocal languages. It does this by foregoing the resources needed to describe a class of good inferences whose goodness relies on mixed disambiguation in certain ambiguous languages. There is nothing in the logic $\models_{\forall WU}$ that 'corresponds' to the good inference from (10i) to (10ii) in the relevant ambiguous setting.

The relation $\models_{\forall WU}$ can also in-principle undergenerate in a similar way because of the fact that it quantifies over disambiguations. It might be simpler to see the issue by first considering general 'non-logical' entailment relations. Consider an inference corresponding to the sentence in (11).

(11) Rashid is at the bank ∴ Rashid is near a river or lake.

This is a good inference if "bank" refers to the edge of a river or lake. It is not a good inference if "bank" refers to a financial institution. Suppose we investigated general entailment relations by quantifying over disambiguations: a transition between ambiguous sentence types corresponds to a good inference if on all ways of disambiguating ambiguous terms it results in an entailment. On this construal the ambiguous type (11) cannot correspond to an entailment because one of its disambiguations is not truth-preserving. Thus there is a good entailment, lurking in one of (11)'s disambiguations, that is never reflected anywhere in the characterization of entailment over ambiguous types.

We just saw how we undergenerate entailments when focusing on uniform disambiguations by missing out on logical relations among different ambiguous terms on particular mixed disambiguations. Here, by quantifying over disambiguations, we undergenerate in a different way: by missing out on potential logical or conceptual relations between one disambiguation of an ambiguous term (like "bank"), and some *non*-ambiguous terms (like "river" or "lake"). By requiring inferential connections to exist on all disambiguations, we overlook the cases where one particular disambiguation (even a uniform one) is doing logical or conceptual work.

This can happen in the logical setting as well, though it is easiest to see if we treat a language where logical terms of some kind are themselves ambiguous. For example, suppose we are working with an epistemic logic where "knows" is treated as a logical term (so the interpretation of that orthographic type as

we interpret logical relations must be 'fixed' up to the resolution of ambiguities). But suppose that while this word can express knowledge, there is also a co-spelled word "knows" such that "A knows that ϕ" expresses the claim that A denies ϕ. Now an intuitive principle of epistemic logic (allowing quantification into propositional position) that captures the factivity of what *we* would univocally express with "knows", namely (12), cannot be a logical principle any longer.

(12) For all x and p, if x knows that p, then p.

This is because there are some disambiguations of this sentence on which it is false: people do not actually (let alone necessarily) deny only true things. Of course, it's still the case that knowledge (as we would put it) is factive. It's just that with an ambiguity in "knows", and having foregone the resources to disambiguate it in a logic like $\forall$WU, we have foregone the resources to give one disambiguation of (12) rather than another a privileged logical status.

These two points about undergeneration are connected with a broader lesson about what kind of 'indirect' information about inference $\models_{\forall\text{WU}}$—or really any entailment relation which foregoes information about actual disambiguations—could be providing us with. Let's go back to (9).

(9) Rashid is at the bank ∴ Rashid is at the bank.

There are at least two good inferences reportable by (9) (alongside at least two bad ones), which roughly correspond to (9a) and (9b).

(9) (a) Rashid is at the banking institution ∴ Rashid is at the banking institution.

(b) Rashid is at the edge of a river or lake ∴ Rashid is at the edge of a river or lake.

Here is an important question: would a transition between *ambiguous* sentences in (9) in a 'logic' considering it to be a good transition represent only the good inference in (9a), only the good inference in (9b), or could it represent some *third* inference type over and above those given by the disambiguations in (9a) and (9b)? It should be clear that it does not *merely* model (9a) or (9b), by consideration of symmetry. But it is also important to acknowledge that it does not model some third inference over and above them (somehow mysteriously brought into being, or maybe simply brought to light, by the presence

of the relevant orthographic contingency). What this means is that if there is a 'logical' property that belongs to (9) in virtue of quantifying over disambiguations, it is not one that belongs to any single inference. In particular, when a language conventionally associates homonymous or co-spelled expressions with different contributions to propositional content, *it does not thereby open up new avenues for reasoning or deduction, or give the linguistic resources for characterizing new reasoning or deductions*. No convention of spelling or pronunciation could have such surprising effects.

Now, why would this matter for the study of inference? After all, I've granted that one can abstract from any kinds of linguistic properties when developing a logic that one wants, as long as one finds the class of inferences that result from the abstraction of sufficient interest. So why not think the relation among ambiguous types in (9) could help characterize a class of inferences *including both* (9a) and (9b), even if it represents no further inference by itself?

The answer is that it *can* do this. The concern is that, at least sometimes, abstracting from disambiguations can generate gerrymandered classes of inference, with no clear interest *beyond* their connections to conventions of orthography or pronunciation. When this happens, classes of inferences picked out in this way do not reveal interesting *bases* for inferring in the way other logical relations tend to do.

Let me try to bring out this problem a little more clearly by considering an idealized example. Focus for the moment on the unary truth-functional operators in a bivalent setting, of which there are four.

p	$\cdot p$
t	t
f	f

p	$\neg p$
t	f
f	t

p	$\uparrow p$
t	t
f	t

p	$\downarrow p$
t	f
f	f

Since these tables exhaust the possible unary connectives for the bivalent setting, the following chart yields all possible inferences between such connectives. (A checkmark in the leftmost four columns indicates an entailment from the row value to the column value, and a checkmark in the rightmost four columns indicates an entailment from the column value to the row value.)

	$\Rightarrow$				$\Leftarrow$			
	$\cdot p$	$\neg p$	$\uparrow p$	$\downarrow p$	$\cdot p$	$\neg p$	$\uparrow p$	$\downarrow p$
$\cdot p$	✓	✗	✓	✗	✓	✗	✗	✓
$\neg p$	✗	✓	✓	✗	✗	✓	✗	✓
$\uparrow p$	✗	✗	✓	✗	✓	✓	✓	✓
$\downarrow p$	✓	✓	✓	✓	✗	✗	✗	✓

Consider now a language for propositional logic with a symbol $\circ$ added that is ambiguous between the two connectives $\neg$ and $\uparrow$. Then its inferential connections to unary connectives, in a logic like $\models_{\forall \mathrm{WU}}$, would be as follows (where there are inferential successes only for columns where we find successes for both $\neg$ and $\uparrow$) .

	$\Rightarrow$				$\Leftarrow$			
	$\cdot p$	$\neg p$	$\uparrow p$	$\downarrow p$	$\cdot p$	$\neg p$	$\uparrow p$	$\downarrow p$
$\circ p$	✗	✗	✓	✗	✗	✓	✗	✓

Note that although $\circ$ can *only* ever be used to express a unary truth-functional connective by assumption, nonetheless its inferential behavior in a logic like $\models_{\forall \mathrm{WU}}$ corresponds to that of *no possible* unary truth-functional connective. (That is, its inferential profile given by the entire row matches that of no single unary truth-functional connective.)

Now, in the context of our hypothetical language for the ambiguous $\circ$, $\models_{\forall \mathrm{WU}}$ does specify a class of inferences involving that term. For example, the validation of inferences of the form $\circ\phi \models_{\forall \mathrm{WU}} \uparrow \phi$ characterize the classes of inference from $\neg\phi/\uparrow \phi$ to $\uparrow \phi$. And $\neg\phi \models_{\forall \mathrm{WU}} \circ\phi$ and $\downarrow \phi \models_{\forall \mathrm{WU}} \circ\phi$ characterize the classes of inference from $\neg\phi$ to $\neg\phi/\uparrow \phi$ and $\downarrow \phi$ to $\neg\phi/\uparrow \phi$ respectively.

But imagine being someone who uses a univocal language containing at least one expression for each of the four possible unary connectives and trying to understand what is interesting about the classes of inferences just specified. There is *no* way of specifying those classes of inferences in the univocal language in logical terms (that is, by giving them a common 'form'), while respecting the fact that all inferences in these classes are between unary truth-functional connectives. Any way of specifying the class would have to relate sentences containing unary truth-functional connectives, with the semantics of those unary connectives held constant as a logical matter. But once this is

done the pattern of inferential connections corresponding to $\circ$ can't be obtained.

What this means is that to understand what *unites* inferences involving $\circ$ into classes by $\models_{\forall WU}$, one *must* consider them as the products ambiguity. That is, what unites the inferences into classes in the logic—the 'form' that they share—is that they would be grouped together in certain ways were orthographic conventions to fail to distinguish between certain connectives. The class is thus *conceptualized* by reference to orthography. It is a class of inferences that is made interesting because of the fact that one could have written $\neg$ and $\uparrow$ using the same symbol. What is more, because of this, this class of inferences doesn't reveal any interesting *basis* for making an inference. Even for those who speak the ambiguous language containing $\circ$, the grounds for performing inferences that would otherwise be expressed with $\neg$ and $\uparrow$ remain the same as they always were. As we've just stressed, expressing two concepts with the same sign doesn't somehow generate new grounds for old inferences, nor does it provide the resources for inferences of a new type.

It is worth emphasizing that this is not true of other classes of inferences traditionally picked out in a logic, even if this is done via schematization. Suppose we consider a propositional logic for a language with negation and disjunction in the bivalent setting, and consider the validation of disjunctive syllogism, given schematically as: $\phi \vee \psi, \neg\psi \models \phi$. This schema singles out a class of inferences from premises to conclusions that differ along many dimensions (e.g., how much syntactic complexity there is in the premises, etc.). What unites the inferences is that they share a 'form' which can be the ground or basis for the goodness of the inferences—a form exhibited by the class. Importantly, the way the class is picked out has nothing to do with the contingencies of orthography. Exactly the same class of inferences can be formalized in logics for a range of languages such that the orthography for disjunction or negation varies between those languages. As long as one can express the logical operations in some way or other, one can see what is interesting about this class. So the class is not conceptualized by reference to orthography. And in part because the orthography one choses to express concepts in one's language plays no important role in singling out the class, it can reveal a *basis* for good inference: a non-linguisticized mode of inferring a conclusion from premises on account of their form that could be recognized and shared regardless of how one wrote out the sentences that expressed its components.

So if one is concerned with inferential relations, although logics for ambiguities can in *some* sense pick out classes of inference, they do so in a gerrymandered way that can make an interest in the class partially parasitic upon an interest in contingencies of orthography or pronunciation. One could certainly have such an interest in principle. But I suspect this is quite far removed from the concerns of most logicians. And it would certainly represent a large departure from the interests of those using logic to investigate the bases for good inference.[42]

I should acknowledge that there is a second reason one could be interested in the classes of inferences picked out in a logic where ambiguities persist. One could be interested in specifically *metalinguistic inferences*—inferences whose subject matter is the sentences of the ambiguous language. That two sentences in a language containing ○ are related by $\models_{\forall WU}$ allows us to infer from the claim that the premise *sentence* is true on a uniform disambiguation of ○, along with all true claims about the syntax and semantics of the language presupposed in the consequence relation, that the sentence acting as conclusion is also true on that same uniform disambiguation. The sentences related by logical consequence, of course, needn't be about language at all. They might concern the weather, mathematics, or politics. In this sense there has been an even more radical departure from traditional deductive inferential logic which is only interested in language as a *tool* for specifying classes of inferences that needn't have anything to do with language. On the new way of understanding how $\models_{\forall WU}$ studies inference, the class of inferences is again conceptualized by reference to contingent orthography, now in an even more straightforward way. But we are at least given information about *bases* for various inferences. It is just that these are now bases for inferences *only* about the semantic properties of sentences of a particular ambiguous language, or perhaps some narrow class of such languages.

In this way logics for ambiguities, though possible, are potentially quite abnormal. It might be worth contrasting these reasons for marginalizing logics for ambiguity from other such considerations, which concern the relative

[42] I've only argued that logics for ambiguity *sometimes* fail to give us information about good inferential bases. But isn't this compatible with their *sometimes* grouping inferences together in ways that do reveal useful bases? Might it not even occur for (9) above? The answer is that of course the logic may group according to proper bases. The problem is that nothing in the logic *distinguishes* between these two kinds of classes—it is 'indifferent' between them. Accordingly nothing in the logic *tells* us when we have a good basis. It has to be antecedently known.

ease or difficulty in resolving ambiguities. David Lewis, when exploring the grounds one could have to interpret and endorse certain relevance logics, came up with the idea that one could be led to some logics of this kind in the face of persistent, unresolvable ambiguities. When he considered *why* these logics for ambiguity could be of interest he had the following to say.

> Logic for ambiguity—who needs it? I reply: pessimists.
>
> We teach logic students to beware of fallacies of equivocation. ... The recommended remedy [for such fallacies] is to make sure that everything is fully disambiguated before one applies the methods of logic.
>
> The pessimist might well complain that this remedy is a counsel of perfection, unattainable in practice. He might say: ambiguity does not stop with a few scattered pairs of unrelated homonyms. It includes all sorts of semantic indeterminacy, open texture, vagueness, and what-not, and these pervade all of our language. Ambiguity is everywhere. There is no unambiguous language for us to use in disambiguating the ambiguous language. So never, or hardly ever, do we disambiguate anything fully. So we cannot escape fallacies of equivocation by disambiguating everything. Let us rather escape them by weakening our logic so that it tolerates ambiguity ...
>
> (Lewis, 1982, 439–40)

Lewis qualifies this by noting that in many contexts, even if full disambiguation were impossible, as long as partial disambiguation is possible (so that we get truth, or falsity, on all remaining 'compatible' disambiguations) there is no *need* for a logic of ambiguity. We can get by with partial disambiguations that create enough uniformity to return us to (say) the classical setting. Still, he concedes, if even partial disambiguation of this kind is impossible, there may be space for logics of ambiguity to gain in importance.

Lewis begins by asking who *needs* a logic for ambiguity. This is an intriguing formulation since 'need' is a bit of a strong criterion for making a logic worthy of investigation. Who 'needs' a propositional logic, when there are first-order logics responsive to additional quantificational structure that propositional logics ignore? And then who needs those first-order logics when one can simply investigate general entailment relations? What I've been arguing in

this section is that even on weakened standards of 'mere interest,' rather than dire need, there are grounds to be suspicious of the importance of logics for ambiguity.

Still I think Lewis is right, in asking where we might be *forced* to consider logics for ambiguity, to raise the specter of pessimism about disambiguation as a conceptual possibility. But even in this context it would be important to highlight the change in focus that the transition to logics of ambiguity would bring. It is not merely that we are investigating the old relations from the non-ambiguous setting with somewhat fewer resources. Now we are forced, by our inability to disambiguate and arrive at the contents of the sentences related by a logic, to a metalinguistic study of relations among the sentences themselves. And in so doing we can lose sight of genuine inferential bases for the contents expressed by sentences of the language, and in these cases must either abandon the search for such bases, or ask instead about bases for inferences solely *about* the language.

For these reasons I feel it is clearest to say that *there is no deductive inferential logic for unresolved strong ambiguities*, even for the starkest pessimist. Pessimism about disambiguation should lead to pessimism about the point of a deductive inferential logic (even if we can study formal relations among ambiguous sentences for other purposes). A logic for deduction must disambiguate wherever it can. And where it can't, there is an important sense in which no true logic of deductive inference is possible.

10.5 Ambiguity and Context-Sensitivity

Consider now the following view about context-sensitivity.

> Context-sensitivity is, as far as deductive inferential logic is concerned, more or less like strong ambiguity. It just happens to be a conventionalized, rule-governed ambiguity. Granted, context may be able to fix the meaning of an expression by different means than those that would typically resolve a lexical ambiguity. For example, perhaps lexical ambiguity is typically resolved by speaker intentions, whereas context-sensitivity can be resolved by broader features of the setting of a speech act. And obviously the contents of context sensitive expressions are often closely related to each other in important ways (e.g. with gradable adjectives)

> which strong lexical ambiguities (as opposed to instances of polysemy) need not be. But there is no clear reason a logic of deduction would be concerned with these nuances. Accordingly, a deductive inferential logic for context-sensitive terms should treat them like ambiguities: just as ambiguity should routinely be banished from such logics, so too should context-sensitivity. One can of course investigate settings in which it is not banished. But this should come with the acknowledgment that one has shifted away from investigating good deductive inference *through* language, and has begun instead to investigate metalinguistic information about a particular language's idiosyncratic connections to classes of inferences, metalinguistic inferences specific to that language, or something disconnected from inference entirely.

Though I don't endorse all aspects of this picture, I think it is a natural starting point for thinking about the role of context-sensitivity for deductive inferential logic. And it is striking how far removed it is for the logic for context-sensitive expressions that Kaplan proposed with LD. When Kaplan develops a logic for context-sensitivity, he does not allow the logic to 'see' how context-sensitive expression have their denotations resolved. Instead he *universally quantifies* over contexts (within the range given by 'proper' contexts). And, up to some caveats to be discussed shortly, he *holds fixed* the contributions of context throughout a sentence (which naturally suggests a view that would hold such contributions fixed across sentences when assessing a consequence relation). That is, he opts for a logic of context-sensitivity that mirrors the choices of the logic for ambiguity $\models_{\forall WU}$ from §10.4.

What is more, Kaplan's methodology, his rhetoric, and even some of his theoretical choices, strongly suggest that his choices represent a *privileged*, if not an exclusive, way to investigate the logic of context-sensitivity. This is especially striking given that the critical focus of Kaplan's logical apparatus is a context-free sentence type or a character—which is the analog in the case of lexical ambiguity of an ambiguous sentence type. For example, recall that Kaplan says of these sentence types or characters that they are the proper objects of logical distinctions, unlike the contents of such sentences (at least when character and content come apart):

> E. Corollary 3 *The bearers of logical truth and contingency are different entities. It is the character (or, the sentence, if you prefer) that is logically true, producing a true content in every context. But it is the content (the proposition, if you will) that is contingent or necessary.*
>
> (Kaplan, 1989b, 539)

Additionally, Kaplan treats characters, not contents (at least where they come apart), as the bearers of epistemic properties like apriority:

> [The features giving rise to the logical truth of "I exist" and "something exists"] correspond to two kinds of a priori knowledge regarding the actual-world…
>
> (Kaplan, 1989a, 597)[43]

And finally Kaplan suggests that characters, unlike contents (at least when they come apart), should be equated with the cognitive significance of a mental episode.

> …a character may be likened to a manner of presentation of a content. This suggests that we identify objects of thought with content and the cognitive significance of such objects with characters.
>
> E. Principle 1 *Objects of thought (Thoughts) = Contents*
>
> E. Principle 2 *Cognitive significance of a Thoughts = Character*
>
> (Kaplan, 1989b, 530)

In this section, building on the work from §10.2–10.4, I want to apply pressure to all of these aspects of the Kaplanian picture. In particular, I will argue for the following two batches of claims (parts of which will already be familiar).

(I) In a logic for deductive inference for perspectival thought, it is not only reasonable to quantify over 'contexts' and to hold their contributions constant, but *necessary* to do so. But this is because such 'contexts' are

[43] See also the suggestion at Kaplan (1989b, 538), that true demonstratives give rise to contingent a priori truths like those advanced by Kripke (1980).

no longer *linguistic* contexts. Once one abandons the goal of modeling perspectival thought, then *even* for perspectival context-sensitive terms it is often perfectly possible, and revealing, to resolve contextual contributions, and allow them to vary within one or more sentences, in assessing semantic relations among sentence types that have a broadly logical character.

(II) When one considers *non*-perspectival context-sensitive language, quantifying over contexts and holding their contributions fixed looks, up to one small caveat, just like doing so in the case of ambiguity. For example, it intuitively undergenerates along the two dimensions discussed in §10.4. What is more, all of Kaplan's claims about the *epistemological* status of linguistic character (that embodies cognitive significance and that it is the object of the a priori/a posteriori distinction) are implausible in the context of non-perspectival context-sensitive language. It is possible to restore some plausibility to these claims if we acknowledge that in quantifying over linguistic contexts, logical inquiry becomes an inherently *metalinguistic* enterprise furnishing us with heavily conditionalized a priori knowledge about language and speech act situations.

Let me take these points in turn.

When I sketched the logic LD^* for *de se* inference, I followed Kaplan both in quantifying over formal objects I called "contexts", and in keeping contextual contributions uniform (e.g., each instance of "I" was to be resolved in the same way throughout a sentence, and even between premises and a conclusion). But in this setting I was abandoning the use of contexts a means of modeling linguistic contexts—circumstances in which a speech act could take place. And I was no longer presuming that terms like "I", "now", etc. had their customary linguistic meanings. This is most apparent from the fact that I allowed 'contexts' in this sense to be agent-free. Contexts for LD^*, on the interpretation I gave them, are merely modeling *values for the parameters relative to which an instance of* de se *cognition could be evaluated for correctness*, and words like "I", "now", etc. were simply used to mark the places where these parameters engaged with *de se* information characterized by other components of the sentence. And given that interpretation, quantifying over 'contexts' and holding their contributions fixed is not only natural, but obligatory. The motivation for quantifying over these parameters and holding them fixed is pre-

cisely that for quantifying over a world parameter and holding that value fixed across premises and a conclusion when evaluating the goodness of ordinary truth-conditional inferences. It is no count against an inference from "Biden is president" to "Biden is president" that the first is true at one world where Biden is president, and then false relative to a different world where he is not. What we care about in an inference is preservation of correctness relative to a fixed world when evaluating premise and conclusion. And we don't care about such preservation for only one world, but a range of worlds, as this is what secures a property of 'safety' that makes an inference good. This requires us to quantify over the relevant worlds. If *de se* attitudes have their correctness relativized to agents and times in addition to worlds, then we must quantify over these and hold them constant when assessing good inference as well.

What is the justification for quantifying over contexts and holding their contributions fixed when we return to the view of contexts construed as modeling speech act situations? Kaplan does not, as I read him, say much about either issue. I think he felt that treating character as the object of logical properties by quantifying over contexts gave intuitive verdicts for the perspectival context-sensitive terms of interest to him (and on this, I would largely agree). And, as we've seen several times now, he did have the following to say about why he relativized truth to a sentence/context pair, rather than an actual speech act.

> …it is important to distinguish an *utterance* from a *sentence-in-a-context*. The former notion is from the theory of speech acts, the latter from semantics. Utterances take time, and utterances of distinct sentences cannot be simultaneous (i.e. in the same context). But to develop a logic of demonstratives it seems most natural to be able to evaluate several premises and a conclusion all in the same context. …We do not want arguments involving indexicals to become valid simply because there is no possible context in which all the premises are uttered, and thus no possible context in which all are uttered falsely.
>
> (Kaplan, 1989b, 522)

This is a strong justification for not assigning truth or logical properties to actual speech acts. But it is no justification for making the contextual contributions relative to which sentence are evaluated constant. Indeed, when Kaplan

considers perspectival context-sensitive terms whose extension easily shifts intrasententially, he modifies his semantics to allow partial resolutions of contextual information, and for contributions from context to shift within a sentence.

To understand this, we need to give some more of the details of Kaplan's treatment of what he called "true demonstratives", which are indexicals which "require, in order to determine their referents, an associated demonstration."[44] These include expressions like "this", "that", or "you", as well as deictic uses of personal pronouns like "he"/"she"/"it". Demonstratives without a demonstration are, as Kaplan puts it, 'incomplete.' This incompleteness is to be distinguished from mere referential failure. A demonstrative may have an associated demonstration but fail to refer because (e.g.) the speaker is hallucinating the object at which they are pointing. But this failure is distinct from the kind of failure when the demonstration itself is simply absent.

It is common for demonstratives to occur multiple times in a sentence accompanied by distinct completing demonstrations. For example, there is a use of (13) on which it which it seems to express an instance of the transitivity of identity.

(13) If that is that, and that is that, then that is that.

Kaplan actually explored two different formal treatments of demonstrative expressions, each of which allowed for (13) to express an instance of the transitivity of identity. On the first treatment, which Kaplan called the "Corrected Fregean Theory", a demonstrative like "that" effectively contributes a rigidifying operator *dthat* to the logical form of sentences. And a demonstration supplies a demonstration type δ which, in the formal language, is essentially just a definite description (that may contain indexicals like "I" or "now"). (13), then, could have the following form.

(13′) If $dthat[\delta_1]$ is $dthat[\delta_2]$, and $dthat[\delta_2]$ is $dthat[\delta_3]$, then $dthat[\delta_1]$ is $dthat[\delta_3]$

Another treatment Kaplan considered was to introduce subscripts on demonstrative terms in the language handled by the logic. He then refined contexts to include sequences of *demonstrata*—the objects (if any) picked out by the demonstrations that accompanied the respective demonstratives at the relevant context of utterance. On this view, (13) would have the form in (13″).

[44] Kaplan (1989b, 490).

(13″) If that_1 is that_2, and that_2 is that_3, then that_1 is that_3

There are further options to explore here. For example, Braun (1996) considers a view on which context-sensitive terms are context-shifting devices. On this view, one use of a context-sensitive term in a context c can shift a new context c' so that a new use of the same term can, even intrasententially, receive a different denotation from c'. And Braun explores, and endorses, yet another view on which "that" has as its linguistic meaning a function from demonstrations to characters.

The differences between these formal approaches matters less to me than the following fact about all of them: on each, the definition of truth-at-a-context is being fed partial information about a context (construed as a possible speech act situation) to avoid equivocation.

This point can sometimes be obscured by the formalism. To clear this up, let me introduce some added terminology. Call a *speech act context* a metaphysically possible situation for speech—something like a part of a metaphysically possible world. Call a *formal context* the formal object in our semantic theory that we use to model speech act contexts.

The first thing to note is that demonstrations that complete demonstratives are parts of speech act contexts. They are metaphysically possible events. Early on, Kaplan considers demonstrations to be, very roughly, something like acts of pointing.[45] Later he considers them to be directing intentions of an agent.[46] Doubtless there are further options and refinements to consider. But the point remains that these are metaphysically possible occurrences of some kind.

The second thing to note is that Kaplan regards it as part of the meaning of a demonstrative that it latches onto demonstrations in this sense. Kaplan writes "the meaning of a demonstrative requires that each syntactic occurrence be associated with a directing intention."[47] The importance of this fact becomes clearest in "Afterthoughts." In that document, Kaplan clarifies that because demonstrations need to be completed by aspects of speech act contexts, modeling speech act contexts with formal contexts that simply *guarantee* the existence of a demonstration can be misleading. Recall Kaplan's distinction between an utterance—the notion which belongs to speech act theory—and

[45] Kaplan (1989b, 489–91).
[46] Kaplan (1989a, 582–4).
[47] Kaplan (1989a, 587).

an occurrence—which is an abstract pairing of a sentence and a context. Kaplan says:

> On my current view, the referent of a true demonstrative is determined by the utterer's intention. But if occurrences don't require utterances, how can we be sure that the requisite intention exists in every possible context? We can't!
>
> (Kaplan, 1989a, 585)

What Kaplan is saying here is that speech act contexts need not contain the demonstrations that completed demonstratives require. Note that this is not an idiosyncratic feature of Kaplan's choice to treat demonstrations as intentions. It would be true on *any* conception of a demonstration that integrates them into possible speech act situations.

Because of this, Kaplan suggests that the definition of validity as truth-in-all-proper-contexts is actually inadequate to the treatment of true demonstratives. After all, if we quantify over proper (formal) contexts, and these formal contexts adequately model speech act contexts, then there will always be many such contexts without demonstrations to complete demonstratives, rendering them defective. The 'logic' of true demonstratives would be trivialized.

Instead, Kaplan suggests that for formal contexts to properly model speech act contexts, they should contain sequences of entities that mark three possibilities: a demonstratum (where demonstration that successfully demonstrates completes a demonstrative), a null element (where a demonstration that fails demonstrate completes a demonstrative), and a marker of *inadequacy* (where no demonstration corresponds to the demonstrative). The logic for true demonstratives then quantifies not over proper contexts, but a subset of these called 'appropriate' contexts, where requisite demonstrations for all demonstratives are present.[48]

It's worth noting that Kaplan considers, and rejects, the idea that we should stipulate away the problems here by divorcing formal contexts from speech act contexts—as Kaplan puts it, by "[imposing] an intention on the agent whether he has it or not."[49] Kaplan notes, among other worries, that this

[48] Kaplan (1989b, 585-6). As far as I can tell, this definition makes the test for validity *sentence-relative*. This unusual aspect of the redefinition of validity could have striking implications for the resulting logic. For example, it threatens to generate valid conjunctions that lack valid conjuncts; see Appendix C for a discussion.

[49] Kaplan (1989a, 586).

is in danger of making it so that "impossibilities come out true" if we impute to the agent an intention, or a collection of intentions, that it is metaphysically impossible for them to have.

I think Kaplan is right to view the demonstrations that complete demonstratives as components of speech act contexts (though I am neutral on what components they are). But it is worth noting that on this assumption any non-trivial logic for demonstratives, and especially one which can model the use of (13) to express an instance of the transitivity of identity, is then one which allows logic 'access' to *some* 'disambiguating' information from speech act contexts.

(13) If that is that, and that is that, then that is that.

(13″) If that_1 is that_2, and that_2 is that_3, then that_1 is that_3

For example, (13″) can be valid for Kaplan. But when it is, this is because we are focusing only on formal contexts that model speech act contexts where there are three separate demonstrations, and the logic is helping itself to information about which such demonstrations correspond *as given in the speech act context* to which instances of the demonstrative "that" in (13). This is a noteworthy amount of contextual information that we have integrated into the validity claim.[50]

Note that it is not obvious that we *have* to do any of this. There is another 'logic' for true demonstratives that does not incorporate any information from speech act contexts—the one that defines validity by reference to all proper contexts. It is just a trivializing one. This raises a number of questions:

(a) Are there any reasons, beyond its utility, to think the logic that integrates *some* information from speech act contexts for true demonstratives is privileged vis-à-vis the trivializing logic?

[50] Braun (1996) distinguishes a third tier of meaning above character and content for demonstratives: one which takes as input a demonstration to yield a character. With this distinction one might maintain: even when we give logic access to information about demonstrations, logic still only governs character. I would be fine with this characterization, as long as one acknowledges the implications of a terminological reshuffle. Now it is simply the case that thing one calls the 'character' of a demonstrative or the sentence containing it has had partial information about some speech act contexts fed into it. And logic now governs *this* kind of object.

(b) If it is merely utility that justifies integration of information from speech act contexts in our logic, why stop at merely *some* of that information? Why not allow for a range of logics that integrate more?

(c) Once we integrate some information from speech act contexts for demonstratives in our logic, why not integrate similar information for *other* perspectival 'pure' indexicals (especially temporal indexicals like "now" or "today")?

Kaplan makes some remarks about questions (a) and (c), which we can consider now. (We'll come back to (b) shortly thereafter when we come to consider non-perspectival context-sensitivity.)

Here is Kaplan, addressing the question of why we *must* modify our semantics to cope with the peculiarities of demonstratives, but are not similarly forced to do so for "now" or "today".

> Why do we not need distinct symbols to represent different syntactic occurrences of "today"? If we speak slowly enough (or start just before midnight), a repetition of "today" will refer to a different day. But this is only because the context has changed. It is a mere technicality that utterances take time, a technicality that we avoid by studying expressions-in-a-context, and one that might also be avoided by tricks like writing it out ahead of time and then presenting it all at once. It is no part of the *meaning* of "today" that multiple syntactic occurrences must be associated with different contexts. In contrast, the meaning of a demonstrative requires that each syntactic occurrence be associated with a directing intention, several of which may be simultaneous. And if it happened to be true that we never held more than one such intention simultaneously, *that* would be the mere technicality. …
>
> The basic fact here is that although we must face life one *day* at a time, we are not condemned to perceive or direct our attention to one *object* at a time. …
>
> Thus within the formal syntax we must have not one demonstrative "you", but a sequence of demonstratives, "you_1", "you_2", etc., and within the formal semantics the context must supply not a single addressee, but a sequence of addressees, some of

> which may be 'null' and all but a finite number of which would presumably be marked *inappropriate*.
>
> (Kaplan, 1989a, 586–7, footnote suppressed)

Kaplan here appeals to the idea that it is part of the very meaning of a demonstrative like "that" or "you" that it corresponds with an associated demonstration—a changing feature or component of a speech act context. *If* true, this *might* be explain why we have to distinguish among various demonstratives in a sentence by indexing them to demonstrations. But it would not be any explanation of why we must quantify over anything less than the full range of proper contexts. That could only be justified on the basis of the utility of feeding more contextual information into the logic.

It is also worth noting that once this is acknowledged, even if Kaplan might be right that we 'must' distinguish between occurrences of "that" in the syntax handled by our logic, and that we need not do so for "today", that is not really the pertinent way of framing the issue. The question is whether we *can usefully* distinguish between contextual contributions for expressions like "today". As far as I can tell, Kaplan supplies no reason to think that is not true. And modest reflection seems to indicate the opposite.

For example, it is easy to index uses of "today" or "now", and to refine formal contexts to include sequences of times relative to which instances of these word types could be interpreted. One might, following Kaplan's idea that we face life one day (or second) at a time, require that if time-sensitive expression e_1 occurs before e_2 in a sentence, then these must correspond to times t_1 and t_2 respectively such that t_1 is at least as early as t_2. (And perhaps even more complexity might be needed for an argument that had multiple sentences as premises alongside a conclusion.) But actually it is not obvious that any of this needs to be done. If it is metaphysically possible to time travel, or time is cyclical, it may be that no combination of time assignments should be ruled out as a logical matter. In fact, the possibility of time travel generates temporal analogs to Frege puzzles, just like those Kaplan supplied for demonstratives ("that is that" pointing at two parts of an object that is partially occluded). Suppose I have a wand that sends me back in time a few seconds. How many seconds? I'm not sure. As it happens, it takes me five seconds (of 'personal time'[51]) between utterances of "today" in (14). So I utter it, beginning at 11:59:59pm, careful to tap the wand just before saying the second "today".

[51] See Lewis (1976).

(14) Today is today.

This might be my way of expressing a conjecture that the wand is sending me back at least five seconds in time. We could model my utterance in a logic for indexicals with (14′).

(14′) Today$_1$ is today$_2$.

If it happens that my wand sends me back exactly eight seconds, the content of each "today" ('in context(s)') would be the exactly same. But I might not recognize this. There could be value in a logic that did not make my assertion correspond to a validity. For example, how many seconds my particular time-travel wand sends me back in time does not seem to be an a priori matter.

Of course, it is not obviously possible to justify shifting values for an actuality operator, or for the first-person pronoun. (Though the latter case raises some interesting questions. Is it possible for others to (as it is colloquially put) 'finish my sentences'? If so, two instances of "I" in a 'single sentence' might be said correspond to different speakers.[52]) But even if these indexicals cannot take on different values, it is simply because, necessarily, the parts of a single sentence *could* not, compatible with their linguistic rules, pick up different contextual contributions vis-à-vis world or speaker within a single sentence. So we would have a principled exception for them which would not apply to temporal indexicals.

So the answer to question (a) above is that there seems to be nothing beyond theoretical interest that privileges the resolution of some amount of contextual information as a logic checks for validity in the presence of true demonstratives. And, given this, the answer to (c) is that there are perfectly legitimate reasons to explore an extension of Kaplan's partial integration of contextual information for temporal indexicals like "now" or "today".

Note again the contrast with a logic modeling *de se* deductive inference, where the choices for the logic are obligatory, and fixed by the subject matter. We've just seen that it is possible, even though the system is uninteresting, to investigate a logic for the *linguistics* of demonstratives that floats free of *any* contextual information. But with respect to demonstrative *thought* this is not the case. A demonstrative thought—a thought about an object from a perspective—is one whose accuracy continues to be assessed with respect to

[52] One might even get a Frege puzzle using the first-person pronoun. "Am I …" Ted begins asking. The time-traveler in front of him finishes: "…me? Yes."

a center (essentially, a world, time, and agent) and no more. Demonstrative thoughts from within a perspective can differ based on how, within a perspective, one singles out an object to think about. But *the way* of singling out an object is not a further feature relative to which the correctness of the thought is assessed. This is why a logic for demonstrative thought would *have to* resolve information about referential focus that would correspond to a demonstration: failure to do so really would just lead to equivocation in the logic, and the abandonment of the concern with inference. One can actually see this division of labor in Kaplan's early treatment of the form of a demonstration.

> …it does seem to me to be essential to a demonstration that it present its demonstrata from some perspective, that is, as the individual that looks thusly *from here now* …We now have a kind of standard form for demonstrations:
>
> The individual that has appearance A from here now.
>
> (Kaplan, 1989b, 525–6)

When we 'disambiguate' a demonstration in perspectival thought, we resolve features that individuate the mental state type. What is being disambiguated is, e.g., the value for A. We cannot disambiguate further than this—by resolving the time or place—because the correctness of the thought expressed by a demonstrative is perspective-relative. So its correctness conditions are misrepresented if it is not allowed to vary in its correctness with respect to agent and time. Kaplan is loath to include the disambiguated information as part of the *linguistic content* (the proposition) expressed with use of a demonstrative (and, for all I know, he is right to do so). But in a logic for *de se* deduction one should *embrace* the integration of the disambiguation directly within the information relative to which correctness is assessed. When we are concerned with perspectival thought, it is clear that this information is part of what individuates the relevant mental state types.

Even the relativity of time in the course of a temporally extended *de se* thought does not give freedom in how to model deduction. This is clearest on the property account of *de se* cognition favored by Lewis, on which contents of *de se* attitudes are given by properties of world-bound agents. E.g., when I wonder whether (as I might express it) "now is now" as I tap a time-traveling wand, I wonder a property—the property of being an agent who was sent back

in time exactly the amount of time it took for them to utter a given sentence (or think a given thought). I am right if I have this property, wrong if I don't. To model the case I *need* to relativize the correctness of my attitude at least to a property. But I do not need to, and indeed cannot, relativize the correctness of what I wonder *beyond* that property. The idea that the property relative to which a *de se* thought was assessed for accuracy would shift has not yet been given any sense.

So when modeling features of language (as Kaplan for the most part purported to do), we have great discretion. We are free to make the formal choices Kaplan did with respect to the logic of demonstratives and temporal indexicals. But we are also perfectly free to model in other ways. We could be more context-neutral in our specification of the logic for demonstratives than Kaplan was, or feed more contextual information into a logic for temporal indexicals than he did, and in neither case would we stray outside the bounds of the 'properly logical.' In modeling thought, we have no such discretion, and are forced into exactly the modeling choices Kaplan made.

What about (b)?

(b) If it is merely utility that justifies integration of information from speech act contexts in our logic, why stop at merely *some* of that information? Why not allow for a range of logics that integrate more?

The answer to this question is best seen by turning to consider *non*-perspectival context-sensitive expressions. In this domain, we often see a shifting set of contributions by context just like with the case of demonstratives.[53] Consider the case of gradable adjectives like "clever" or "large". Intuitively, the things that count as clever or large are those satisfying a sufficiently high standard along some scale or metric (of intelligence or size) where both the standard and the metric are subject to influence from linguistic context. For example, consider a sentence like (15) said by an entomologist to a graduate student when the latter uncovers a subpopulation of wood lice that are uncommonly good at solving mazes.

(15) If you are clever, you'll see why you should put the lice that are clever through maze A and the dunces through maze B.

Intuitively, the scales and standards that would be used to evaluate the cleverness of a graduate student are quite different from those that would be used to

[53] Cf. Crimmins (1995).

evaluate cleverness for wood lice. In principle, this kind of intrasentential shift of contextual contributions is always possible. And these shifts can be present alongside intuitively logical connections as in (16).

(16) If every large house contains a large dog, and every large dog has a large flea on it, then every large house contains something with a large flea on it.

An utterance of (16) would be one where linguistic context rapidly shifts the standards for size: what it takes for a dog to be large involves a weaker standard than for a house to be large, and for a flea to be large involves standards that are weaker still.[54] If one were to model the contextual shifts in a language as we did those for demonstratives, we would index adjectives as follows.[55]

(16′) If every large$_1$ house contains a large$_2$ dog, and every large$_2$ dog has a large$_3$ flea on it, then every large$_1$ house contains something with a large$_3$ flea on it.

Imagine yourself in a context where you speak (16) or hear it so that it clearly expresses the disambiguation in (16′). Is what is said or what is thought—the *disambiguated content*—assessable for logical properties? This seems hard to deny. Surely someone who has asserted (16) in a context where it is interpreted along the lines of (16′) could justify the truth of what they said on logical grounds, were they challenged; they would be right to believe what they said on broadly familiar logical grounds; and so on.

Kaplan's logic, applied uniformly to context-sensitive expressions, would not treat the relevant disambiguation of (16) as a logical truth. First, it is not even *assessable* for logical properties (it belongs to the realm of content, not character). What is more, although the character of (16), from which the content of the disambiguation is derived, could count as a logical truth, it would be a logical truth in virtue of quantifying over contexts that hold contextual contributions fixed. So the logicality of (16′) is nowhere reflected in the logical properties ascribed to characters.

[54] One might think the use of the adjective in modifying a nominal creates a compositional, obligatory effect that crowds out any influence of context. But there are reasons to think that the effect here is neither compositional nor obligatory (see Kennedy (2007)). Even if it were, the general point being made could still stand on the basis of *some* example of intrasentential contextual shift.

[55] We could, of course, introduce *much more* form to model the special semantic features of gradable adjectives, but this matters little to the conceptual points I am trying to make.

What we seem to find here is a relatively straightforward analog to the problem for the logic of ambiguity $\models_{\forall WU}$ that arose because it required disambiguations to be uniform. We saw that, intuitively, logical connections can arise from relations between mixed disambiguations. What we are seeing here is that, again intuitively, logical connections can arise from relations between mixed *contextual* disambiguations. A logic that foregoes the resources to model those connections appears to undergenerate in its aim of understanding how context-sensitivity impinges on logical relations, roughly as does a logic for ambiguity.

Kaplan also ascribes logical and epistemic properties to the analogs of ambiguous sentences. Ambiguous sentences intuitively have no deductive logical properties independently of their disambiguations (except, as we've seen, on idiosyncratic reorientations of the purpose of logic). What about context-free sentences? Does (16) *devoid* of disambiguation—that is *free* of any way of construing what it takes to be large—have a logical status of any kind? Could the context-free sentence, or the character it represents, be a priori? Does it correspond to a mode of presentation—a distinctive mode of thought? Though the situation here isn't quite as bad as with ambiguity, I think an affirmative answer to these questions hardly feels straightforward. In particular, the idea that abstract linguistic rules corresponding to a gradable adjective like "large" give one's *way of thinking* of largeness is much less intuitive than the idea that the abstract linguistic rules corresponding to a perspectival indexical correspond to distinctive mode of thought about an object. We'll return to discuss this issue in more depth momentarily.

Recall that just as one problem for $\models_{\forall WU}$ arose because it held disambiguations fixed, another arose because it quantified over disambiguations. This hid important logical connections between some disambiguations and some *non*-ambiguous terms. There appears to be an analog of this problem for logics of context-sensitivity as well. This is especially apparent with respect to quantification and modality.

The dominant view of natural language quantification treats the domains of quantifiers as sensitive to context.[56] Theorists disagree in some measure about how this context-sensitivity is regulated—for example, whether or not it is mediated by some syntactic element present natural language logical form. But many share the view that however the influence of context is mediated,

[56] See, e.g., von Fintel (1994), Stanley & Szabó (2000).

an English sentence like (17) can be used to express different contents loosely glossed by specifications like (17b–17e).

(17) (a) All the beer is in the fridge.
(b) All the beer *I bought at the store* is in the fridge.
(c) All the beer *I bought at the store, except that I dropped on the way home* is in the fridge.
(d) All the beer *gifted to us by our neighbors* is in the fridge.
(e) All the beer *that Alisha likes* is in the fridge.

Suppose that quantification behaves in this way. What does a logic for this kind of quantifier look like, in the sense Kaplan gives? It would be one which checks for truth, or truth-preservation, across *all* specifications of quantifier domains. This is an almost absurdly general approach to thinking about the logic of the quantifier. Consider a simple quantificational inference involving what would appear to involve Universal Instantiation. I'm teaching a class and you ask how your friend, Marta, is doing. I might make the following claims, and inference.

(18) Everyone in the class passed
Marta is in the class
∴ Marta passed

This feels like a logically secure inference. But if the quantifier in (18) is sensitive to context, the only subject of *logical* properties, on the Kaplanian model, is the character of my utterance. And truth preservation in this instance is not secured by character, as it is not secured on *every* way of resolving the context. There are contexts (not the one in which I spoke of course) in which Marta is not in the domain of the quantifier (even if other students in my class are in the domain, and they, along with Marta, passed).

In fact, a logic attributing validity only to characters would effectively invalidate every instance of universal to particular inferences (whether for abstracted unary, or more realistic binary quantifiers) for context-sensitive quantifiers in which the conclusion could not be independently established.

$$\frac{(\forall x)(Fx)}{Fa} \qquad \frac{\begin{array}{l}(\forall x : Fx)(Gx)\\ Fa\end{array}}{Ga}$$

Granted, if there is a syntactic element in logical form mediating the involvement of context, these inferences will look a little different—roughly as follows[57]—and the grounds for *some* failures of the inference would be more transparent.

$$\frac{(\forall x)(f_i(x) \rightarrow Fx)}{Fa} \qquad \frac{\begin{array}{l}(\forall x : f_i(x) \wedge Fx)(Gx) \\ Fa\end{array}}{Ga}$$

But this is not fully relevant to the current point. Everyone should agree that if we can at least apply logical properties at the level of content, then there must be some logically invalid instances of inferences corresponding to English sentences like (18). The question is whether we are willing to embrace a characterization of consequence on which *no* representation of (non-trivial instances of) universal instantiation for context-sensitive quantifiers *ever* counts as valid, simply because there are always possible contexts which exclude any given element from the domain of quantification.

It is worth noting that when logicians encounter the context-sensitivity of quantification, they are perfectly happy to resolve at least *some* of the influence of context before beginning to attribute logical properties to quantificational claims, and deductions using them. If someone were to model the inference corresponding to (17) using a first-order language, they would adopt a standard semantics for the first-order regimentation that treats "Marta" as a constant symbol. The structure of a first-order model would then do two things: (i) it would ensure that this constant symbol receives a denotation from within the universe of the model and (ii) it would ensure that quantifiers range over that universe—i.e. over a domain that includes that denotation. I've already discussed the importance of the first assumption in Chapter 7. But the second assumption is easy to overlook. It is possible to create models (not a model of the familiar sort of course) with a domain of existents from which the denotations of constants are drawn, but to restrict the values of quantifiers to a subset of that domain. Logicians and others who acknowledge the context-sensitivity of ordinary language quantification, and model inferences expressed with those quantifiers using first-order models, are making a substantive choice about what kinds of objects are the subject of logical properties. In particular, they ascribe logical properties to *partial contextual disambiguations*, and they do so without compunction.

[57] See Stanley & Szabó (2000).

These kinds of intuitively logical connections between particular contextual disambiguations and context-insensitive terms can also be uncovered in the treatment of modality. It is familiar that natural language modals like "must" or "can" can often be used to express multiple different *kinds* of modal relations. For example, "must" can be used to express a form of epistemic necessity (as in (19a)), deontic necessity (as in (19b)) or teleological necessity (as in (19c)), among other forms.

(19) (a) Your keys aren't here. You must have left them in the drawer.

(b) You promised. You must go see her.

(c) You want to win? You must train harder.

Given the pervasiveness of similar variation cross-linguistically, the received view of natural language modality follows Kratzer (1981) in holding that a term like "must" is not ambiguous, but rather that any variation in the 'flavor' of modality "must" expresses traces to the influence of linguistic context.

The exact details of how context produces this effect needn't concern us here. Consider the use of the modal in (19a). Does it follow from the claim that you must have left the keys in the drawer that you *did* leave them in the drawer? This does seem to follow. "Must" on its epistemic reading appears to be factive.[58] But consider (19b). Does it follow from the claim that you must fulfill your promise that you will? Of course not. It is possible, as a conceptual matter, for people to fail to fulfill their obligations.

These connections strike me, as they have struck many logicians, as logical ones. But if we follow Kaplan's definitions, and modal 'flavor' is sensitive to context, factivity (for example) is not a logical property of epistemic "must". This is because epistemic "must" is "must" *fed* a certain amount of contextual information—the information from context that resolves modal flavor. If the logic of "must" is the logic belonging to its character, in the Kaplanian sense, then the only question about the factivity of "must" that we can raise is whether the truth of "must *p*" secures the truth of "p" in virtue of its character. And it does not.

Intuitively, both universal instantiation and factivity are logical connections that belong to some applications of quantifiers and modals respectively. And in that respect, a definition which attributes logical properties only to

[58] Well, at least it appears that way. There is some controversy here (see, e.g., von Fintel & Gillies (2010)), but it doesn't matter for the conceptual point I want to make here.

character appears to suffer from the same problems as did $\models_{\forall WU}$ for quantifying over disambiguations.

With respect to quantifiers and modals, the problems for attributing logical and especially epistemic properties to context-free sentences or characters also grows. Imagining thinking what is expressed with the help of the quantifier in (18), in its relevant context. Does the context-neutral semantics for its quantifier (on which it has *no* determinate domain of quantification associated with it[59]) give something like a 'mode of presentation' of that content, in the sense in which this seemed so plausible for an indexical like "I"? When one thinks what is expressed by the modal locution in (19a), does the context-neutral interpretation of that modal (on which it has *no* modal flavor—epistemic, deontic, or otherwise) give something like a 'mode of presentation' of its content in context? My sense is that if it is even *possible* to retrieve such an abstract way of thinking about quantification or modality, it seems completely absent from the relevant episodes of thinking corresponding to the uses of the sentences in context. Indeed, it was a striking claim of Kratzer's that there *was* a schematic object, neutral on the flavor of modality, corresponding to the modal "must" to begin with. If there is a flavor-neutral 'mode of cognition' that accompanies every use of a modal, Kratzer's claim should have been trivial, and hardly something that required interesting empirical justification.

When we encountered problems of undergeneration in the context of ambiguity, I suggested that this was symptomatic of a more general problem. This was that the 'logic' had lost touch with the aim of logical inquiry as studying inference with language playing a merely instrumental role. Instead, a logic foregoing resources to disambiguate begins to take language itself as its object of study, at least in part. Is that true in the case of context-sensitivity as well? As it happens, matters here are more subtle. I think there is much more room in the domain of context-sensitivity, than in the presence of lexical ambiguity, to abstract away from some features of context and recover good inferential forms—that is, good bases for inference. This is my main caveat for the view given at the start of this section about the parallels between logics for ambiguity and logics for context-sensitivity.

Now, some lessons about inferential bases do carry over from ambiguity to

[59] Importantly, this is distinct from an *unrestricted* reading of the quantifier—as that would merely be another (trivial) restriction.

context-sensitivity. Consider a simple inference involving a non-perspectival context-sensitive term.

(20) (a) A is big $\therefore$ A is big.

(b) A is big_1 $\therefore$ A is big_1.

(c) A is big_2 $\therefore$ A is big_2.

(d) A is big_1 $\therefore$ A is big_2.

(e) A is big_2 $\therefore$ A is big_1.

… …

The context-free type (20a) corresponds to many good inferences represented by (20b–c), and bad (or potentially bad) inferences represented by (20d–e). Just as in the case of an 'inference' from an ambiguous sentence to itself, I am inclined to regard the complete disambiguations (the complete resolutions of context) as representing the proper objects of inference. And I would maintain that there is not some *further* inference over and above these corresponding to the context-free type (20a), at least in the non-perspectival case.[60]

Still, we can characterize such inferences, grouping them together using context-free types and even types fed partial information from context. In the case of ambiguity, I suggested this kind of maneuver embroiled us in the study of orthography. But here, I think it is less clear that we are embroiled in a study of language—even of context. We can say some plausible things about how we continue to care, directly, about general rules for thought. Consider the first-order case. There, I noted that a first-order model effectively feeds contextual information to any context-sensitive quantifiers by ensuring that named objects figure in quantifier domains. That fact is held fixed if one quantifies over models in assessing quantificational validities and consequences. I think there *is* a good inference type corresponding to the rule of Universal Instantiation that results. It is not merely that from a quantifier bearing a *specific* domain of quantification containing a named object one can infer facts about that object. Rather, I suspect it is possible to engage in an inference corresponding to Universal Instantiation while 'ignoring' or 'indifferent' to the details of the domain, beyond that it contains the object about which one infers. If so, there *would* be a good inference type here helpfully characterized through a partial, but not full, resolution of context. I think something similar could be said

[60] I will speculate more on the perspectival case in §10.6.

about the factivity of epistemic modals. In fact, there is good reason to think that the logical property of factivity holds constant, even as the worlds over which epistemic modals quantify continue to vary dramatically from context to context. And arguably one can infer on the basis of a general factivity, rather than from the understanding that factivity is witnessed for a *particular* contextual modal domain. If so, again, we'd have use for a partial resolution of context in classifying a reasonable *basis* for an inference.

The key difference between context-sensitivity and strong ambiguity is that there are *semantic ties* between full contextual resolutions of a single context-sensitive expression that don't arise on full disambiguations of strong ambiguous terms, essentially by the stipulations on membership in the latter class. And these semantic ties among contextual resolutions can seemingly track helpful commonalities between good inferences—commonalities that can even register bases for good inference. It is mere orthographic happenstance that financial institutions and river edges are designated by homophones in English. But it is no orthographic coincidence that families of height scales are characterized by a single expression like "tall" in English. And thus is it no wonder that we could exploit the connections between the various contextual resolutions of context-sensitive terms like "tall" to uncover important inferential commonalities.

So we can make the following generalizations: A concern with deduction should generally drive a logic toward full resolution of strong ambiguities. But that concern can instead drive us to consider *various* levels of resolution of context-sensitivity (including fully, and not at all) for logics in the presence of context-sensitivity. Which of these variations properly maintain contact with investigations of sound bases for good deductive inference may be a complex matter that needs to be taken on a case-by-case basis. After all, sometimes a logic of this form *can* lose its grip on possible inferential bases—a completely context-neutral logic for natural language modals would be an example.

In spite of this concession about the utility of context-neutral (or partially context-neutral) sentences in logical inquiry, however, we should continue to be dubious of Kaplan's use of character as the object of logical distinctions, as the occupiers of a role of cognitive significance, and as the subject of epistemic properties like apriority.

There is perhaps one way of safeguarding this role for character: by embracing the claim that logic is in fact simply studying language and linguis-

tic relations. After all, one *can* reconstrue the purpose of a logic for context-sensitivity in such linguisticized terms, just as we saw possible for logics of ambiguity. We saw that a logic for ambiguous terms can be understood a providing heavily conditionalized a priori truths about particular languages, or language types. A logic for context-sensitivity could do this as well. What is more, the class of languages it would tell us about would be much larger than that for ambiguity, because of the non-accidental ways in which contents are linked in disambiguations of context-sensitive terms.

In fact, when Kaplan talks about the sense in which logical truths are a priori, sometimes he casts the apriorities in precisely these kinds of heavily linguisticized ways. Consider a continuation of a quotation above about the a priori status of "I exist" and "something exists" in *LD*.[61]

> [The features giving rise to the logical truth of "I exist" and "something exists"] correspond to two kinds of a priori knowledge regarding the actual-world…Corresponding to the first feature, there is our knowledge that certain *sentences* always express a truth regarding the world in which they are expressed. Corresponding to the second feature, there is our knowledge that certain *facts* always hold at a world containing a context.
>
> (Kaplan, 1989a, 597)

The forms of knowledge corresponding to the logical truths are overtly linguisticized. The first kind of a priori knowledge is straightforwardly about language. What one knows a priori explicitly concerns sentences. Even if we switched to talk of a priori knowledge about the characters of these sentences (which we could well do), this merely becomes knowledge about linguistic rules and their applications. And though Kaplan lays emphasis on "facts" in describing his second kind of a priori knowledge, he might better have laid stress on "context"—a *speech act* context. We know, a priori, that if there is someone around to possibly speak, something exists. We should be happy to grant that this kind of conditionalized a priori knowledge exists. But I think we should be suspicious that it, or anything corresponding to it, is of much *general* interest.

Let me wrap up with one final issue. Kaplan said that the objects of logical properties like validity were characters, unlike contents, which were the proper objects of attributions of necessity. We have a partial vindication of this claim

[61] On validity of the latter, see n.6.

in the sphere of non-perspectival context-sensitivity. We *can* sometimes attribute logical properties to context-neutral sentences *or* sentences paired with partial resolutions of context. Either way, these objects are not truth-evaluable, and so (on any view) should not be the proper objects of necessity claims. Still, it is important to recognize that this does not have the effect of conceptually divorcing logical notions from notions of necessity. It would not, for example, vindicate the claims we saw Gillian Russell make about the significance of Kaplan's work. On the contrary, properly construed, the logic again simply reinforces the connections between validity and necessity. This is because the importance of assigning logical properties to (partially) context-neutral sentences *derives* from their ability to characterize classes of good inference (where context is *fully* resolved) and where we can see each member of the class counts as a good inference precisely for preserving truth at all metaphysically possible worlds. That is, the importance of assigning the logical properties, at least in the non-perspectival cases, derives from the ability of these assignments to characterize classes of metaphysical necessities, or classes of transitions that are metaphysically necessarily truth-preserving. *All characterized inferences involving quantifiers, modals, gradable adjectives, and so on transparently preserve truth at all metaphysically possible worlds*.

An analogy might help clarify the point. We noted before that logicians sometimes attribute logical properties like validity to *schemas* which can be instantiated in various ways. In this context we could say, as some do,[62] that logical properties can belong to these objects—schemas. *These* objects of validity are of course distinct from the bearers of properties like necessity. After all, the schemas themselves don't express anything true or false, let alone anything necessarily true or false. This is all fine and good. But it would be *highly* misleading to go on to say that this showed that logical truths need not be necessary, contrary to popular assumption. It is obvious that the importance of attributing logical properties to the schemas can derive from the necessary statuses of their instances. I've argued that in the first-order case, instances of first-order schemas are all necessary (or necessarily truth-preserving), and it is at least in part because of this that the schemas are of any interest to begin with. To all appearances, the very same kind of connection that holds between schemas and their instantiations also holds between any not-fully-disambiguated non-perspectival context-sensitive sentences and their full disambiguations.

[62] Again Quine (1970/86, 50-1).

10.6 Final Thoughts

We've covered a *lot* of ground in this chapter, so it is worth stepping back to collect the claims I've advanced and chart some of their relations. The goal of this section has been to try to understand how a deductive inferential logic would adapt to the presence of each of perspectival thought, context-sensitivity, and ambiguity as it influences the study of good deduce inference through language. The main positive contentions have been as follows.

(i) A deductive inferential logic must eventually broach the non-linguistic issue of how to model perspectival or *de se* cognition, and the inferential relations among *de se* cognitive states.

(ii) On an exceptionalist treatment of the *de se*, there is a body of information associated with a *de se* cognitive state relevant to intuitively good inference and whose correctness is relative to a (metaphysically possible) world/time/agent triple, or a property. A logic for inference modeling perspectival thought would naturally integrate (perhaps ad hoc) expressions in a formal language to mark the place of parameters relative to which the information connected with a *de se* cognitive state is relativized. Once this is done, in assessing conditions on good inference by tracking relations of correctness preservation, the logic *must* quantify over the values of those parameters and hold them constant.

(iii) The resulting logic for perspectival thought would resemble LD^*, which is a slight modification of Kaplan's Logic of Demonstratives LD. The *interpretation* of this logic would be quite unlike that Kaplan supplied for LD. Notably, the formal analog of contexts would not correspond to contexts of utterance or linguistic contexts of any kind.

(iv) This logic would invalidate some of LD's suspicious existence entailments, and also reveal that logical truths and consequences for perspectival thought are undergirded by a natural generalization of metaphysical necessity to perspectival bodies of information.

(v) A general logic for *linguistic context-sensitivity*, in contrast with one for perspectival thought, need not resemble LD, especially when non-perspectival context-sensitivity is taken into account. Kaplan's choices

to treat logical objects as context-neutral, to hold contextual contributions fixed, and to quantify over contexts when assessing logical relations are all to some extent arbitrary ones in the purely linguistic setting. Indeed, each of those choices can often undergenerate in the task of modeling good inference, and can deviate into a study of language itself, rather than a study of inference *through* language. All of these problems have direct analogs in logics which forego resources to disambiguate ordinary lexical ambiguities.

(vi) Still, a logic for deduction in the presence of context-sensitivity, *can* sometimes justifiably abstract from some, and perhaps even occasionally all, disambiguating information—as is much less commonly justifiable for ordinary lexical ambiguities. This can be done as a means of grouping inferences into classes that reveal important bases for good deductive inference. These bases typically trace to semantic commonalities among various contextual resolutions of a context-sensitive term that secure their contributions to (metaphysically) necessary truth-preserving transitions.

(vii) This means we can sometimes take an object that is context-neutral or partially contextually resolved and attribute logical properties to it. In this way, as Kaplan suggested, validity could belong to objects that are not plausible bearers of necessity. But, at least focusing on non-perspectival context-sensitive terms, these logical properties *derive* from the goodness of the inferences in the class characterized by it: those inferences expressed by contents supplied on 'full' contextual resolutions. Not only are these inferences good in virtue of necessarily preserving truth, but it is clear that the importance of the logical attributions stem from this fact.

Kaplan did not characterize his logic for context-sensitive terms as a logic for deduction. He hardly mentioned reasoning or deduction in his discussion. And he developed a logic specifically for perspectival context-sensitive language, not for other forms of context-sensitive expressions. So there is a way in which few of the above claims *need* come into conflict with Kaplan's logic or his claims about it. Still, I think that asking about the form that logics for deduction would take in the presence of non-perspectival linguistic context-sensitivity and perspectival thought uncovers some concerns about Kaplan's

choices in developing LD and his glosses on its significance.

Above I was careful to separate out two logics, or two classes of logic: one for perspectival thought and one for linguistic context-sensitivity. Once these are separated out, there is a concern that the shape of Kaplan's logic, and some of its appeal, derive from a periodic conflation of aspects of perspectival thought with the linguistic conventions that accompany their expression. Kaplan clearly sees these as connected, with a result that both linguisticizes mentality, and mentalizes linguistic rules.

The core locus of this conflation would occur for the Kaplanian notion of character. Characters are defined as ways of modeling *linguistic* rules or conventions. But Kaplan imbues these characters with certain forms of epistemic significance that don't intuitively belong to any linguistic rules. He slots them into the role of cognitive significance, treats them as an object of apriority, and clearly leans on both of these ideas in treating characters as the objects of logical properties like validity.

By "Afterthoughts," Kaplan began to acknowledge the unusual nature of using a linguistic device in this way. He says he "follow[ed] Frege in using a strictly semantical concept (character), needed for other semantical purposes, to try to capture [the] idea of cognitive value."[63] He qualifies this to some extent. In particular, he notes that if names are not context-sensitive then the characters of coreferring names like "Cicero" and "Tully" appear to have identical characters.[64] He concludes that "[s]ince it is indisputable that distinct proper names have distinct cognitive values, the project of discriminating cognitive values of proper names by character is immediately defeated."[65] This tells us that not *all* differences of cognitive value trace to differences of character, so that character could not embody cognitive significance generally. But Kaplan was clearly tempted to see continued connections. In a footnote, he says:

> Even granting that we cannot *articulate* the rules of character for all directly referring expressions, we may still recognize a difference in cognitive value when presented with a pair of terms of different character. There may still be a correlation between distinct characters and distinct cognitive values. Jospeh Almog suggests

[63] Kaplan (1989a, 598)

[64] Of course, Kaplan already recognized this point in Kaplan (1989b, 562). The point is that only in "Afterthoughts" do we find reflections on how we should view the limitations this imposes on the 'cognitive significance' role of character.

[65] Kaplan (1989a, 598).

> that we might express the point by saying that cognitive value *supervenes* on character.
>
> Kaplan (1989a, 597-8,n.67)

But even this correlation or supervenience claim does not hold. Consider that there are variations between languages among which modal 'flavors' (metaphysical, epistemic, etc.) are licensed by particular modal lexical items. This variation can result (in principle, if not in fact) in the Kaplanian characters of natural language modals differing even though they can each be used to express some particular familiar modal flavor—say, epistemic modality. Though uses of these modals to express epistemic modality would diverge in Kaplanian character, it strikes me as implausible to claim that these uses *must* come with a corresponding change in cognitive significance, at least in the senses in which Kaplan is clearly interested. As I noted when discussing Kratzer's work, speakers can be oblivious to the ways in which the modals they use *could* have been used to express other modal flavors.

Linguistic characters for *non*-perspectival context-sensitive expressions are not tied to, nor even generally correlated with, some species of cognitive significance. In the non-perspectival case, it is implausible to claim that characters are the objects of apriority (unless one means that they figure in conditionalized a priori knowledge about linguistic rules). And finally their ties to logic are underwritten by the necessity of the contents they express in context. Why would we not say the same thing of *perspectival* context-sensitive terms like those Kaplan made his focus?

This leaves us with several options. A first option is to say that there is a sharp divide between the behavior of perspectival and non-perspectival context-sensitivity and try to maintain that the epistemological and logical status of perspectival linguistic character reaches a status wholly other than that of non-perspectival character.

A second option is to unify the logical treatment of perspectival and non-perspectival context-sensitivity by making both *about* language. We can make logic concerned with a priori conditionalized truths about language or linguistic rules. Indeed, when Kaplan explicitly articulated the status of the apriority of logic he seemed to frame the apriorities in explicitly linguisticized terms. It is worth emphasizing, of course, that this treatment of logic is far from forced. *Even* for non-perspectival context-sensitive expressions, we've seen we can maintain a familiar and properly *instrumental* concern with language in

studying forms of reasoning.

There is of course a third option that becomes much more natural when we see that a logic for perspectival thought can be developed independently of consideration of the language of context-sensitivity. On this view, there are certain very loose parallels between perspectival linguistic characters and perspectival thoughts. But the parallels do not warrant treating perspectival context-sensitive terms in any way differently than their non-perspectival counterparts. The parallels mostly create *hazards* for conflation, not opportunities for unified explanation. On this view, Kaplan fell into a natural trap.

If we see Kaplan's logic as the result of mental/linguistic conflations, much begins to make sense. A logic for perspectival cognition must quantify over something like parameters for agent, time, and world, and hold these constant, just as Kaplan does for linguistic contexts. A logic for demonstrative perspectival cognition would have to disambiguate between 'perspectival demonstrations,' but also would have to stop short of further 'disambiguation' on pain of misrepresentation of its subject matter, again just as Kaplan does for linguistic demonstratives. The parameter-neutral objects of logical distinctions for perspectival logic also occupy a role of cognitive significance, and can be bearers of apriority (or a similar epistemically privileged property), just as Kaplan claimed for characters. Indeed, they make up a subject matter to be modeled directly in a core conception of logic—the one that I have been developing in this book—that concerns itself with the conditions on good deductive inference.

It is worth noting that to take Kaplan to have conflated aspects of perspectival thought and perspectival talk is *only* to criticize some of the epistemological and logical elaborations he gives of his semantics. It is fully compatible with (and often simply leans upon) the idea that Kaplan's *compositional semantics* for perspectival context-sensitive terms is essentially correct. It is also worth noting that claiming that Kaplan's work results from mental/linguistic conflations comes with further argumentative burdens that I will not be able to take up here. Once we abandon the use of linguistic character in making sense of cognitive significance, we still owe some explanation of how we engage with context-sensitive language when we are ignorant of the features of context which would fix their extensions. This can occur frequently with perspectival context-sensitive language, and it arguably occurs less frequently with non-perspectival context-sensitive terms. Leaning on features of accompany-

ing perspectival cognition may do some of the work here. But I doubt it will do all of it.

Most notably there is this incredibly challenging issue: when we use a term like "I" without knowing our identity, *what are the relationships* between

(a) the linguistic content expressed by the sentence we utter,

(b) the information associated with the perspectival cognitive state type we are typically in that would lead us to utter the sentence, and

(c) the mental content of that cognitive state type (if this is different from the information associated with it in (b))?

The route I have been exploring puts the following constraint on our answer: *if* Kaplan is right about the linguistic, propositional content expressed with the help of "I", (a) and (b) come apart. That is striking, and also leaves a great deal of latitude when filling in the rest of the picture of these relations, not all of which may be plausible.[66] As I say, I won't be able to explore these questions in detail here. I will be content for now to note the ways in which a broader examination of perspectival thought, non-perspectival linguistic context-sensitivity, and ambiguity pressure us to strongly reconsider core Kaplanian theses about the epistemic features of linguistic character—features that undergird his (necessarily non-inferential) conception of logic.

[66] For example, one natural approach to the relations I am describing is by embracing a form of Guise Theory, but in which one maintains a clean separation between mental guises and linguistic character. But another option is to abandon the conception of linguistic propositional content defended by Kaplan (one might for these purposes try to extend some recent work on contents involved in *de se* communication—see, e.g., Ninan (2010a)).

CHAPTER II

Validity for Information-State Logics

This chapter explores the question of how to develop and interpret logics for languages that make use of a shiftable information-state parameter to capture the semantics of expressions that convey subjective states of uncertainty. Information-state semantics for indicative conditionals and epistemic modals will provide the central cases.

I begin in §II.1 by reviewing a number of examples that appear to threaten the validity of Modus Ponens or Reasoning-by-Cases, taking Vann McGee's well-known counterexamples to the former inference rule as a point of departure. I describe how a family of semantics developed in response to the cases, united by the idea that conditional consequents are evaluated relative to a parameter that can be shifted by conditional antecedents. Several distinct characterizations of logical validity have been proposed for the resulting semantics. In the context of our investigation, this naturally raises the question of which, if any, notion captures the relation appropriate to a logic for deductive inference.

In §II.2, I argue that there is no simple answer to this question, as what an information-state logic for deduction looks like depends heavily on an embedding interpretive framework that draws ties between the semantics and the contents of mental states. To justify this claim, I contrast two frameworks given by Seth Yalcin and John MacFarlane respectively. Though each framework makes use of similar compositional treatments of conditionals and modals, I argue that they give rise to strikingly different applications of logical machinery in the context of modeling deductive inference. In particular, the most perspicuous logic of this kind for Yalcin's framework is essentially given by its modal- and conditional-free fragment, with a result that could be as simple as ordinary classical logic. The logic for MacFarlane's system, by contrast, *must* incorporate modalized and conditionalized language, and ac-

cordingly look quite different. Sadly we cannot see exactly how the logic for MacFarlane's system should pan out because of a lacuna arising from a subtle circularity within his account of the information contained in mental states. MacFarlane's case reveals just how tricky it can be to specify the details required to settle questions about a deductive inferential logic within information-state semantics.

With these two frameworks outlined, I turn in §11.3 to contrast two popular definitions of validity for information-state logics sometimes called 'classical' (or 'diagonal') consequence and 'informational' consequence. I note that there is a tendency to conflate a rejection of the former consequence relation with a rejection of a conception of logic as tracking relations of (necessary) truth-preservation. Focusing on the work of Justin Bledin, I argue that this tendency arises from a conceptual confusion. Once the typical application of information-state-sensitive language is taken into account, we see that informational consequence is in fact the most natural extension of the view of logic as concerned with necessary preservation of truth (though this is obscured by the consequence relation's typical formulation).

After a brief return to explore how McGee's counterexamples to Modus Ponens interact with a recent literature on the 'weakness' of belief in §11.4, I conclude in §11.5 by discussing a trend in developing information-state semantics for probability modals that treats attitude states as fundamentally bearing graded, probabilistic structure. I note that we currently have no adequate models of how to reason, let alone infer, with graded mental states and that this interferes with our ability to give any sense to a deductive inferential logic in this context. I propose some first steps in developing an account of inference for graded states, noting that this seems to require quite radical restructuring of mental information. In particular, we seem driven to replicate the structure of 'full', non-graded attitudes toward contents with probabilistic elements in ways that are not obviously consonant with the original motivations for grading states like belief. Even with this structure, several concerns emerge that either obscure the form of a deductive logic for information-state semantics for probability modals or calls into question its tenability. Without space to pursue matters further, I leave the domain of logics for probabilistic discourse as one were increased attention to the nature of deductive inference may be critical, as without such attention it is unclear we can meaningfully speak of a logic of deduction at all.

11.1 Logical Challenges from Conditionals and Modals

McGee (1985) presented a case against Modus Ponens using several related examples, including the following two.

> Opinion polls taken just before the 1980 election showed the Republican Ronald Regan decisively ahead of the Democrat Jimmy Carter, with the other Republican in the race, John Anderson, a distant third. Those apprised of the poll results believed, with good reason:
>
> > If a Republican wins the election, then if it's not Reagan who wins it will be Anderson.
> >
> > A Republican will win the election.
>
> Yet they did not have reason to believe
>
> > If it's not Reagan who wins, it will be Anderson.
>
> I see what looks like a large fish writhing in a fisherman's net a ways off. I believe
>
> > If that creature is a fish, then if it has lungs, it's a lungfish.
>
> That, after all, is what one means by "lungfish." Yet, even though I believe the antecedent of this conditional, I do not conclude
>
> > If that creature has lungs, it's a lungfish.
>
> Lungfishes are rare, oddly shaped, and to my knowledge, appear only in fresh water. It is more likely that, even though it does not look like one, the animal in the net is a porpoise.
>
> (McGee, 1985, 462)

McGee states that such examples show that Modus Ponens is "not an entirely reliable rule of inference," and "not strictly valid."[1]

While McGee's examples are undoubtedly important ones, it is not easy to trace out their implications, least of all for logic. For example, it is not clear that the two premises above can be true while the conclusion is false. McGee above describes the premises as ones that are believed "with good reason." Later he

[1] McGee (1985, 462-3)

says that there is "ample reason to believe" them, that they are "reasonable to believe," and that they are things we "believe very properly." By contrast, the conclusions are such that we do "not have reason to believe them," or that there "is no reason to suppose" them, or that we "should not believe them."[2] As noted in a critical reply by Sinnott-Armstrong et al. (1986), there is a conspicuous absence in McGee's descriptions of his cases of talk of truth or untruth, which is intuitively the terminology needed to directly cast doubt on Modus Ponens.

Could McGee's claims about reasonable belief entail claims about validity? Not obviously, for reasons we reviewed in Chapter 3. On many, if not most, construals of validity, valid inference need not preserve reasonable belief. Strong connections of this kind would threaten principles like (generalized) conjunction introduction via the Preface Paradox, and many other intuitive logical rules when we consider the the kinds of epistemic circumstances Harman marshaled against logico-normative bridge principles.

McGee didn't make explicit claims about the truth-values of his premises and conclusion, nor claims that would entail them. But could he have made such claims? Not obviously—at least not without significant controversy. Sinnott-Armstrong and his co-authors worry that we need substantial defense to show that the conclusions of McGee's instances of Modus Ponens are not true, rather than true but unassertable on pragmatically problematic grounds. I would voice a rather different worry. In each example, given the truth of the non-conditional premise, the embedded conditional in the conditional premise (which also figures as conclusion) must either must have a false antecedent or a true consequent. Either way, it would be controversial to treat the resulting conditional as simply false. This is straightforward for conditionals with true consequents. But also, as is familiar, it is problematic to assert an indicative conditional in a context in which it is apparent that its antecedent is false. One way of accounting for these intuitions is to say that indicative conditionals exhibit truth-value gaps when their antecedents are false. Views like this (depending on how the gaps compositionally project) may end up requiring the first premise of the above examples to be untrue (despite being reasonable to assert, and reasonable to believe, in the context McGee supplies).[3] In fact,

[2] McGee (1985, 462-3).

[3] These accounts face noteworthy obstacles, of course (see §5 of Edgington (2020), e.g.). I tend to think these problems turn on an over-simplistic conception of truth-value gaps (see Shaw (2014)). Either way, my point here is not to defend one theory of conditionals over oth-

we can arguably set issues of projection aside as long as we accommodate an importation principle for the conditional (whereby one can derive "If p and q, r" from "if p, then if q then r")—which McGee appeared to accept.[4] If we treat this inference as truth-preserving, then the truth of McGee's first conditional premises (of the form "if p, then if q then r")) alongside the falsity of the antecedent of the embedded conditional (i.e. "q"), would require a conditional with a false antecedent (namely "if p and q, then r") to be simply true.

Though he didn't defend truth-value judgments for his conditionals, McGee did give the following indirect argument against the idea that truth-preservation for Modus Ponens could be salvaged in the face of his examples.

> Modus ponens is sometimes thought of not as a rule of inference but as a law of semantics, to wit, whenever ⌜ If ϕ then ψ⌝ and ϕ a both true, ψ is true as well. It is not at all obvious what we are to make of this law, since it is not evident what the truth conditions for the English conditional are or even whether it has truth conditions. Still it seems unlikely that, even if we learned the truth conditions for the English conditional, the semantic version of modus ponens would be vindicated. Let us imagine, on the contrary, that some time in the future linguists will determine the truth conditions for the English conditional and prove that modus ponens is truth-preserving. Assuming that basic zoology will not have changed, a future linguist who sees what looks like a large fish writhing in a fisherman's net a ways off will believe, as I believed,
>
> > If that animal is a fish, then if it has lungs it's a lungfish.
> >
> > That animal is a fish.
>
> Suppose he also believes this:
>
> > It is true that, if that animal is a fish, then if it has lungs it's a lungfish.
> >
> > It is true that that animal is a fish.

ers, but merely to stress the presence of controversy. That is, the point is that the truth-values of McGee's conditionals are straightforward *neither* from an intuitive perspective, *nor* from a theoretical one.

[4] McGee (1985, 465 n.5).

> Then he will be able to prove, using the well-established principle of future semantics that modus ponens is truth-preserving:
>
> > It is true that, if that animal has lungs, it is a lungfish.
>
> He will not, however, believe
>
> > If that animal has lungs, it is a lungfish.
>
> any more than I did. Thus our future linguist will be either in the awkward position of believing the premises of the argument without believing that those premises are true, or else in the equally awkward position of not believing the conclusion of the argument even though he does believe that that conclusion is true.
>
> (McGee, 1985)

I cannot see how this argument is any stronger than a corresponding argument that the Preface Paradox must show that generalized conjunction introduction could not be truth-preserving.[5] The parallel argument runs as follows:

> Suppose linguists discover that the transition from $p_1, \ldots, p_n$ to $p_1 \wedge \ldots \wedge p_n$ is truth-preserving. Then a linguist may believe each premise in a preface paradox, and so believe they are true. They will 'be able to prove, using a well-established principle of semantics' that the conjunction is true. But they will not believe the conjunction. So they are in the awkward position of believing the premises without believing the premises are true, or in the equally awkward position of not believing the conclusion of the argument even though they believe that that conclusion is true.

McGee's argument seems to ignore that if is not rational to infer the conclusion from some premises, it will be equally irrational to infer the claim that the conclusion is true from the claim that the premises are true. This is so *even if* the inference is known to be truth-preserving (however this is established). On the conception of logic I favor, the reason for this is that the validity of an inference is not always sufficient grounds to make it, as I stressed in Chapter 3. But I suspect other conceptions of validity would have some analogous

[5] The point is not about probabilities: we could derive parallel arguments from any puzzling epistemic circumstance where it is not rational to infer.

way of exploiting the parallel with the Preface Paradox (or similar confounding epistemic circumstances). I think McGee is entirely right to be cautious about whether conditionals have truth-conditions. But it is not obviously of help in this context to retreat to preservation of reasonable belief as a standard for validity. That is a bad standard *even* in the case where truth-conditions are clearly secured.

None of this is to say that McGee's examples *don't* raise trouble for Modus Ponens. It is rather to say that they cannot do this on their own. They could certainly form an important part of the case against Modus Ponens as part of a broader set of examples motivating a semantics for conditionals, and an associated well-motivated conception of validity, which paired together would show McGee's cases to invalidate the rule. Indeed, another of McGee's contributions is to develop a semantics for the conditional which would explain our judgments in his examples. It will be helpful to sketch this semantics, as it integrates the key formal element whose relevance to logic I want to try to get clearer on in this chapter: a shiftable information-state parameter.

McGee's conditional involves a slight modification of that put forward in Stalnaker (1968). We can grasp the important elements by considering a simple language fragment of McGee's framework consisting only of sentence letters p, q, etc. and a connective $>$ for the indicative conditional. Sentences are assigned truth-values in McGee's proposal relative to four parameters:[6]

- A possible world w,
- a valuation function $\mathcal{I}$ mapping an atomic sentence to a set of worlds (intuitively, those where the sentence is true),
- a selection function f mapping a pair consisting of a set of worlds and a world $\langle S, w\rangle$ to a world w' (where, intuitively, w' is the 'closest' world to w within S), and
- a set of hypotheses Γ, given by a set of sentences.

The selection function f satisfies three constraints, where S and T are sets of worlds (thought of as propositions):

(i) *Success*: $f(S, w) \in S$.

[6] McGee also relativizes truth to an accessibility relation which I will here presume to be universal, and accordingly suppress.

(ii) *CSO*: $f(S, w) = f(T, w)$ iff $f(S, w) \in T$ and $f(T, w) \in S$.

(iii) *Strong Centering*: $f(S, w) = w$ if $w \in S$.

(i) ensures the world selected is among those where the associated proposition in true. (ii) ensures f can be interpreted in terms of a closeness ordering on worlds. And (iii) ensures every world is the closest to itself.

The integration of the final parameter above—the set of hypotheses Γ—is the key change to Stalnaker's theory. This set of hypotheses keeps track of how conditional antecedents provisionally update a body of information relative to which conditional consequents are evaluated. Suppressing parameters for valuation and selection functions, we evaluate sentences as follows (letting $[\![\phi]\!]^{\Gamma} = \{w \mid [\![\phi]\!]^{\Gamma,w} = 1\}$):

(i) If $\bigcap_{\gamma\in\Gamma} [\![\gamma]\!]^{\emptyset} = \emptyset$, $[\![\phi]\!]^{\Gamma,w} = t$; otherwise:

(ii) for atomic ϕ,

$$[\![\phi]\!]^{\Gamma,w} = \begin{cases} t & \text{if } f(\bigcap_{\gamma\in\Gamma} [\![\gamma]\!]^{\emptyset}, w) \in \mathcal{I}(\phi) \\ f & \text{otherwise} \end{cases}$$

(iii) $[\![\phi > \psi]\!]^{\Gamma,w} = [\![\psi]\!]^{\Gamma\cup\{\phi\},w}$

(i) ensures sentences are vacuously true if a hypothesis set is inconsistent. (ii) says an atomic sentence is true under a set of hypotheses just in case it is true at the closest world where all hypotheses are true. (iii) tells us to evaluate a conditional at a world under a set of hypotheses by first updating that set with the antecedent, then checking for the truth of the consequent relative the initial world and the update.

A sentence is *simply true* at a world w (on an interpretation, given a selection function), just in case it is true relative to w under the null hypothesis set $\emptyset$. Note that given strong centering, if $\Gamma = \emptyset$ the right hand side of condition (ii) reduces to the simpler condition: $w \in \mathcal{I}(\phi)$. So an atomic sentence is simply true at a world just in case the world is among those assigned truth to it by the interpretation function. The more complex truth-conditions that non-trivially invoke closeness will only matter for embeddings in conditional consequents.

The theory gives us intuitive resources for modeling how Modus Ponens could fail when a conditional embeds a further conditional in its consequent. In particular, for three atomic sentences p, q, and r, we have

$$q > r \text{ is simply true at } w \text{ iff } [\![r]\!]^{\{q\},w} = t$$

whereas,

$$\begin{aligned} p > (q > r) \text{ is simply true at } w &\text{ iff } [\![q > r]\!]^{\{p\},w} = t \\ &\text{iff } [\![r]\!]^{\{p,q\},w} = t. \end{aligned}$$

So the embedded conditional is true just in case r is true at the closest q-world, whereas the embedding conditional is true just in case r is true at the closest p-and-q world.

Even if p is simply true at the actual world (e.g., a Republican actually wins), and r is true at the closest p-and-q world (Anderson wins at the closest world to actuality where a Republican wins and Reagan loses), it need not follow that r is true at the closest q world (i.e. it needn't follow that Anderson wins at the closest world to actuality where a Reagan loses). p's actual truth is not enough to guarantee, via any constraint we have imposed so far, that the closest p-and-q world to actuality is also the closest q world to actuality. After all, the closest q world may be one which falsifies p.

Note that what does the important work is the embedded conditional in the consequent (as McGee stresses is the key factor that invalidates Modus Ponens). If p and $p > q$ are simply true at a world for sentence letters p and q, then q is also simply true at that world as well. E.g. if p is actually true, then the closest p-world is the actual world by *Strong Centering*. So if $p > q$ is also actually true—which means that q is true at the closest p world—then q must actually be true as well. This reasoning generalizes to cases where the antecedent p is a sentence of arbitrary complexity. So it is complexity in the consequent that matters.

McGee's semantics for indicative conditionals, in its specific details, was not especially influential in the subfield of semantics devoted to their compositional behavior. (Nor do I think McGee intended for it to be—he was

merely exploring one way to consistently accommodate the intuitions about his conditionals.) But the key technical idea behind McGee's semantics has, if anything, become the dominant view. The key idea I am alluding to is that conditional antecedents at least partially function to shift some informational parameter to which the evaluation of some kinds of conditional consequents may be sensitive. McGee's examples provide some evidence that conditionals themselves could be sensitive to an informational parameter of this kind. But another large class of expressions which seem to exhibit some such sensitivity are modals. Indeed, one of the most influential proposals for the semantics of conditionals among linguists, owing to Angelika Kratzer, posits that conditionals have *no function other* than to restrict an explicit or implicit modal quantifier domain.[7]

If conditionals have this effect, we should expect to find interesting logical behavior among conditionals with modals embedded in their consequents. And many, now familiar examples bear this out. Forrester (1984) introduced the puzzle of the gentle murderer, which is an instance of puzzles in deontic logic surrounding 'contrary-to-duty obligation.'[8] Consider (1).

(1) If Cain will kill Abel, Cain ought to kill Abel quickly.

(1) seems to tell us about the relative merits of quick and slow killing. And since quick, painless killings are better than slow ones, (1) can ring true. But it also seems plausible that Cain should not kill Abel. It even seems that he should not kill him quickly. If so, we appear to be confronted with a violation of Modus Tollens: we cannot conclude what Cain will in fact do merely from the truths that quick murder is wrong and quick murder is preferable to slow. One can get a parallel worry for Modus Ponens: if we learn that Cain will kill Abel quickly (and so will kill him), can we really conclude that Cain is doing what he ought to be?

[7] See Kratzer (1977, 1981, 1986). Kratzer's view familiarly takes inspiration from the treatment of interactions between conditionals and adverbial quantifiers in Lewis (1975). Kratzer's view also treats conditional antecedents as having a *syntactic* function like quantifier restrictors, which raises a host of worries about the logical form of conditional statements, and whether there is any sense to attributing a form like Modus Ponens or Modus Tollens to statements within Kratzer's framework. These are important questions, but not ones that I have space to get into here.

[8] See Chisholm (1963) for a precursor in the form of 'Chisholm's Paradox.' I will present Forrester's puzzle here in a different way than he does, to emphasize the logical challenges that are my particular focus.

Dreier (2009) and Kolodny & MacFarlane (2010) independently present problems for reasoning by cases when conditionals embed deontic modals,[9] with the latter introducing the following much-discussed 'miner puzzle' into the literature.[10]

> Ten miners are trapped either in shaft A or in shaft B, but we do not know which. Flood waters threaten to flood the shafts. We have enough sandbags to block one shaft, but not both. If we block one shaft, all the water will go into the other shaft, killing any miners inside it. If we block neither shaft, both shafts will fill halfway with water, and just one miner, the lowest in the shaft, will be killed.

Action	if miners in A	if miners in B
Block shaft A	All saved	All drowned
Block shaft B	All drowned	All saved
Block neither shaft	One drowned	One drowned

(Kolodny & MacFarlane, 2010, 1)

The puzzle is that the plausible claims (2a)–(2c) would imply (2d) if we could reason by cases. But this conclusion seems false.

(2) (a) If the miners are in shaft A, we ought to block shaft A.
(b) If the miners are in shaft B, we ought to block shaft B.
(c) Either the miners are in shaft A, or they are in shaft B.
(d) We ought to block shaft A or we ought to block shaft B.

Kolodny and MacFarlane claim that uses of Modus Ponens, subsumed in the reasoning by cases, are at fault. (In contrast, Dreier took Modus Ponens to be unexceptionable and looked for the fault elsewhere.) They note that a similar puzzle can be constructed with epistemic modal consequents (a point independently appreciated by Cantwell (2008), who criticizes the use of reasoning

[9] Dreier's case actually makes use of the comparative evaluative "better". But the idea is essentially the same.

[10] The puzzle is based on an example given in Parfit (1981), who attributes it to Regan (1980), though in those contexts it is not used to raise a logical puzzle.

by cases in Edgington (1996)). For example, in the context of a murder investigation we might accept (3a)–(3c), while rejecting (3d) (since we accept that the murder might have happened in the morning and also might have happened in the evening).

(3) (a) If the butler did it, the murder must have occurred in the morning.
(b) If the nephew did it, the murder must have occurred in the evening.
(c) Either the butler did it or the nephew did it.
(d) Either the murder must have occurred in the morning, or the murder must have occurred in the evening.

Again, Kolodny and MacFarlane think Modus Ponens is ultimately at fault. Yalcin (2012b) notes that epistemic modals—especially probability modals—embedded in conditional consequents appear to give rise to challenges to Modus Tollens, using the following example.[11]

> An urn contains 100 marbles: a mix of blue and red, big and small.
> The breakdown:
>
> | | blue | red |
> |---|---|---|
> | big | 10 | 30 |
> | small | 50 | 10 |
>
> A marble is selected at random and placed under a cup.
>
> (Yalcin, 2012b, 1001-2)

Yalcin then notes that (4c) "does not intuitively follow" from (4a)–(4b), which both seem rational to accept.

(4) (a) If the marble is big, then it's likely red.
(b) The marble is not likely red.
(c) The marble is not big.

An important class of responses to many of the foregoing examples using modals is to make use of the same basic idea that McGee employed. That is, theorists relativize the evaluation of modalized expressions to a body of information which conditionals conventionally shift—with the result of disrupting

[11] Again Cantwell (2008) gives similar cases.

otherwise sound logical reasoning via principles like Modus Ponens or Modus Tollens. Of course, there are many alternative routes to explore to deal with the puzzles, some of which would have no interesting disruptive effects for logic.[12] There is an interesting array of arguments for and against such positions. As usual, my goal here is not to adjudicate the disputes. Rather, what I want to explore is the question of what we should say about logic, on my sense of logic, *if* these frameworks are roughly on the right track.

One of the fascinating things about information-state semantics is that there are at least two natural ways to generalize logical relations. Very roughly, one of these preserves a property like truth across of range of cases, and another tracks preservation of structural features of bodies of information. Sometimes these conceptions of logic are seen as complementary, other times as rivals. But at most one of them could give the right definitions for logic, in my sense of logic. That is, at most one could track conditions on good inference. So it is worth asking which (if either) does so.

To that end, it will be helpful to see how the two logical notions arise in the semantics of Kolodny and MacFarlane (henceforth K&M) for modals and conditionals, which is representative of the frameworks I want to explore. K&M's compositional semantics consists in a recursive definition of truth at a point of evaluation, where a point of evaluation is given by a pair $\langle w, i\rangle$ of a possible world (representing an epistemic possibility) and an information state (given by a set of such possible worlds). We have the following clauses for informational necessity and possibility modals $\Box_f/\Diamond_f$, of which deontic and epistemic modals are instances, where a selection function f maps information states to information states.

$$[\![\Box_f\phi]\!]^{w,i} = t \Leftrightarrow \forall w' \in f(i) : [\![\phi]\!]^{w',i} = t$$

$$[\![\Diamond_f\phi]\!]^{w,i} = t \Leftrightarrow \exists w' \in f(i) : [\![\phi]\!]^{w',i} = t$$

Context determines whether the selection function f is epistemic (=e) or deontic (= d) in character. An epistemic selection function e maps an information state to the set of worlds that might, as far as the state knows, be actual.

[12] As regards the behavior of epistemic modals, this is actually my preferred view. See the descriptivist alternative explored in Marushak & Shaw (ms./2020). But such views are extremely controversial. What is more, even if they are correct, the tenability of alternative information-state semantics seems like it is in part an empirical question. If that is right, then even if such information-state semantics in fact faced obstacles, it would still be very important to know what we *would* have had to say about logic had the evidence panned out in their favor.

K&M provisionally take this to be the identity function. A deontic selection function maps an information state into the set of worlds that are as deontically ideal as possible given the initial information. Without characterizing deontic selection functions generally, they are assumed to be *realistic* and some are assumed to be *seriously information-dependent* in the following senses.

> A deontic selection function d is *realistic* iff for every information state i, $d(i) \subseteq i$.
>
> A deontic selection function d is *seriously information-dependent* iff for some information states i_1 and $i_2 \subseteq i_1$ there is a world $w \in i_2$ such that $w \in d(i_1)$ but $w \notin d(i_2)$.

A realistic deontic selection function ensures that what ought to be the case is constrained by what the information allows. Serious information-dependence arises in a selection function when increases in information not only can rule out ideal worlds but change rankings of ideality over worlds. For example "a world in which both shafts are left open may be more ideal than one in which shaft A is closed relative to a less informed state, but less ideal relative to a more informed state."[13]

Finally K&M treat the semantics for the conditional, following Kratzer, as a modifier of modals. In particular, they choose to treat the antecedent of an indicative conditional as an operator [if ϕ]. Conditional consequents that are not explicitly modalized are assumed to have an implicit epistemic necessity modal. Intuitively a conditional shifts the information state parameter to 'incorporate' the information given by the modal antecedent. For non-modalized antecedents this would simply be the set of worlds where the antecedent is true. But for modalized antecedents we need a broader notion given as follows.[14]

> An information state i *supports* ϕ, written $i \triangleright \phi$, iff $\forall w \in i : [\![\phi]\!]^{w,i} = t$

A slight hurdle is that we cannot have a conditional consequent evaluate relative to 'the' updated information state supporting an antecedent ϕ, as many distinct information states can sometimes support modalized antecedents. For

[13] Kolodny & MacFarlane (2010, 133).

[14] Kolodny & MacFarlane (2010) use the terminology of ϕ being 'true throughout' an information state. Sometimes this notion is called "acceptance", a term which I must avoid here because of ambiguities it would create further below when discussing acceptance states.

example, $\Diamond_e\phi$ is supported by every non-empty information state with a ϕ-world in it. And we cannot get around the problem of a plurality of supporting states by focusing on the 'largest' one in the following sense, as there may be no unique largest state of this kind either.

> i' is a maximal ϕ-subset of i iff $i' \subseteq i$, $i' \triangleright \phi$, and there is no i'' with $i'' \triangleright \phi$ and $i' \subset i'' \subseteq i$.

K&M give the example of "we ought to block a single shaft" to emphasize the problem. Intuitively this would be supported by an information state where, in every world, the miners are in shaft A and not B. It should also be supported by an information state where, in every world, the miners are in shaft B and not A. But these states have no worlds in common.

To resolve the issue K&M have conditional consequents evaluate relative to *each* largest information substate that supports the information of the antecedent. This gives our final semantics for the conditional.

$$[\![[\text{if}\phi]\psi]\!]^{w,i} = t \Leftrightarrow \text{for each maximal } \phi\text{-subset } i' \text{ of } i\text{: } [\![\psi]\!]^{w,i'} = t$$

K&M explore many virtues of this account, and explain away some of its unusual features. Here, I am only concerned with the logic they take to result from their compositional semantics. To get to such a logic we need to say how to get to validity and consequence from the recursive definition of truth at a point. K&M opt for what is sometimes called 'diagonal validity'—essentially an extension of the Kaplanian definition of validity I critiqued in Chapter 10.[15] Just as Kaplan focused on 'proper contexts,' we can focus on 'proper points of evaluation,' which will be those pairs $\langle w, i\rangle$ such that $w \in i$. This essentially treats the 'information state of a proper context' (if such there be) as one which cannot err. This would make sense if that information subsumed some batch of knowledge, since knowledge is factive. At any rate, this gives us the following definition of consequence.[16]

> $\Gamma \models_d \phi$ iff for every information state i and world $w \in i$: if $\forall\gamma \in \Gamma$, $[\![\gamma]\!]^{w,i} = t$, then $[\![\phi]\!]^{w,i} = t$

[15] K&M simply call this "validity', though mention the terminology I use here in a footnote. For the record, these relations and some subsequent ones we will consider are relations of *general entailment*. But their relevance to parallel logical entailment relations is straightforward.

[16] This is more of a general consequence relation, rather than a logical one. I'll elide the distinction between these in what follows, since it is relatively clear how a transposition to the logical case would go.

It is easy to see how Modus Ponens can fail relative to this conception of consequence. The truth of a conditional [if ϕ]ψ ensures ψ is true however the information that ϕ is incorporated into an information state. But ϕ's *truth* at a proper point does not ensure it is somehow integrated into the information state of that point. As such, ψ could be true relative to shifted points of evaluation in which ϕ's information is incorporated, while still being false relative to an unshifted point of evaluation at which ϕ is true. Note this is, at an abstract level, just what happens in McGee's semantics that invalidated Modus Ponens (and K&M's framework can be applied in a straightforward way to McGee's example—see Kolodny & MacFarlane (2010, 137)).

K&M account for the miners case by saying Modus Ponens fails for conditionals like (2a):

(2) (a) If the miners are in shaft A, we ought to block shaft A.

(2a) is true because relative to an information state in which all miners are in shaft A, the deontically best worlds are block-shaft-A worlds. But, as K&M put it, while "it may in fact be the case that the miners are in shaft A…that would not make it the case that ["We ought to block shaft A"] is true relative to our original information state—the one that includes both worlds where the miners are in shaft A and worlds where they are in shaft B."[17]

In the immediate wake of K&M's work, a number of philosophers contested their treatment of puzzle, in particular suggesting that Modus Ponens was not at fault. Fascinatingly, some philosophers did not always reject K&M's definition of truth at a point (or at least did not always feel this was the *important* aspect of their account for understanding the puzzle). Rather these authors challenged the characterization of *validity* that K&M employed.[18] The rival conceptions appealed to are what are sometimes called *informational* validity or consequence. The informational conception of consequence checks not for preservation of truth at a privileged set of points of evaluation, but rather checks for preservation of informational support.

$$\Gamma \models_i \phi \text{ iff for each information state } i\text{: if } \forall \gamma \in \Gamma\ i \triangleright \gamma, \text{ then } i \triangleright \phi.$$

Modus Ponens is an informationally valid rule. Consider any information state i which supports the minor and major premises of a Modus Ponens inference.

[17] Kolodny & MacFarlane (2010, 138).
[18] See, e.g., Willer (2012), Bledin (2014).

Note that for K&M's conditional to be supported by i, it must be that for all worlds in i, and antecedent-maximal subsets i' of i, the consequent is true at $\langle w, i' \rangle$. But since i by assumption supports the conditional antecedent, i is the only antecedent-maximal subset of i to consider. So for any world $w \in i$, the consequent is true at $\langle w, i \rangle$. But that is for i to accept the consequent. (And just as the diagonal conception invalidates Modus Ponens for McGee's semantics, the informational conception will validate it.)[19]

The terminology of 'informational' consequence comes from Yalcin (2007), who tentatively endorses it as an entailment relation for his semantics for epistemic modals. But (as Yalcin acknowledges) the notion bears a close affinity to logics developed by dynamic semanticists—see especially Veltman (1996), and more recently Willer (2012). It is explicitly defended as a basis for logical consequence in Bledin (2014) (whose work we will return to shortly).

What is especially fascinating is that K&M did not overlook this alternative conception of validity. Indeed, they explicitly characterize an equivalent notion of validity—*quasi-validity*—and reject it as a candidate for the consequence relation.[20] In particular, K&M say it is a *confusion* to think this is a form of logical validity, since it conflates entailment with entailment from *known* premises.

As I've stressed, at most one of these conceptions of validity can be logical in the sense of helping to track good deductive inference. Which, if either, is it?

[19] If Modus Ponens is not at fault what goes wrong in the Miners Puzzle? We have several options, some of which we will review in §11.3.

[20] An inference from Γ to ϕ is *quasi valid*, $\Gamma \models_{qv} \phi$, just in case $\{\Box_e \gamma \mid \gamma \in \Gamma\} \models_d \phi$. Note that for any w and i, $[\![\Box_e \gamma]\!]^{w,i} = t$ iff $i \rhd \gamma$, and that this latter condition is world-independent. Thus we have:

$$\begin{aligned}
\Gamma \models_{qv} \phi &\Leftrightarrow \{\Box_e \gamma \mid \gamma \in \Gamma\} \models_d \phi \\
&\Leftrightarrow \forall i, \forall w \in i : \text{if } \forall \gamma \in \Gamma, [\![\Box_e \gamma]\!]^{w,i} = t, \text{ then } [\![\phi]\!]^{w,i} = t \\
&\Leftrightarrow \forall i, \forall w \in i : \text{ if } \forall \gamma \in \Gamma, i \rhd \gamma, \text{ then } [\![\phi]\!]^{w,i} = t \\
&\Leftrightarrow \forall i : \text{ if } \forall \gamma \in \Gamma, i \rhd \gamma, \text{ then } \forall w \in i : [\![\phi]\!]^{w,i} = t \\
&\Leftrightarrow \forall i : \text{ if } \forall \gamma \in \Gamma, i \rhd \gamma, \text{ then } i \rhd \phi \\
&\Leftrightarrow \Gamma \models_i \phi
\end{aligned}$$

II.2 Two Case Studies

To start answering this last question, I want to delve into the details of two frameworks: that for epistemic modals in Yalcin (2007), and that for epistemic and deontic modals in MacFarlane (2014).

Yalcin and MacFarlane likely stand out to some readers as developers of highly original approaches to the semantics of modal discourse—Yalcin being a chief advocate of an *expressivist* framework for epistemic modality, and MacFarlane being a chief advocate of a form of *relativism* about epistemic and deontic modal discourse. Seeing their names appear might suggest that I am singling them out for treatment because of these unorthodox positions.

But in some important respects, the expressivist and relativist facets of Yalcin and MacFarlane's respective views are orthogonal to the issues that matter to me. Instead, I single out these theorists because they are unique in giving relatively detailed accounts of the *mental states* that arise when we correctly characterize acceptance states with the help of modalized and conditionalized language. Not only do they give semantics for propositional attitude reports that interact with their respective semantics for conditionals and modals, but they devote noteworthy attention to the question of how the information contained in a mental state is structured, and what implications this has for their semantics of information-state-sensitive expressions. These are topics that many theorists—including some discussing the 'proper' logic for information-state semantics—do not broach at all.

The goal of this section is to show that slight differences in how we approach questions about mentality lead to radically different understandings of how to understand good inference in the context of modal and conditional discourse. I'll begin in §II.2.1 by arguing that a framework for mentality explored in Yalcin's work leaves no room for modalized attitudes that could be premises or conclusions of an inference, and as a result the most perspicuous logic in this setting would be unaffected by the addition of modals—with some caveats it could, for example, simply be ordinary classical logic. Then I'll turn to the quite different framework proposed by MacFarlane in §II.2.2. Here I'll argue that in spite of the fact that MacFarlane's compositional semantics is virtually identical to Yalcin's, the proper logic for his framework should look radically different. Unfortunately, MacFarlane does not put us in a position to ascertain some details of that logic, as his discussion of mental states characterized

by modal language leaves some of their structural features underspecified. (In §11.3 we'll learn some rough lessons about how the logic for MacFarlane could look once the circularity is resolved.)

The key lesson of this section lies in the contrast between Yalcin and MacFarlane: our understanding of mentality and the role of information-state-sensitive language in characterizing it matter tremendously to how we understand a deductive logic for that language. Without specifying those details, there is typically *no answer* to the question of what logic for deduction would be 'correct' to pair with a given information-state semantics.

11.2.1 Yalcin on Epistemic Modality

Yalcin (2007) motivates what he calls a 'domain semantics' for modal discourse in part from the infelicity and embedding behavior of sentences like (5), which Yalcin calls "epistemic contradictions."

(5) # It's raining and it might not be raining.

(6) ? It might not be raining and it's raining.

Fascinatingly, the infelicity of epistemic contradictions persists in suppositional environments and conditional antecedents (as in (8a)–(9a))—a fact which distinguishes them from more familiar Moore-paradoxical sentences like (7) (with embeddings in (8b)–(9b)).

(7) # It's raining and I don't know that it's raining.

(8) (a) # If it's raining and it might not be raining, then …
(b) If it's raining and I don't know that it's raining, then …

(9) (a) # Suppose it's raining and it might not be raining.
(b) Suppose it's raining and I don't know that it's raining.

Yalcin argues that there is some difficulty capturing this data on a standard 'relational' semantics for epistemic modals. These arguments need not concern us here. Instead, let's jump straight to giving a pared down version of the domain semantics Yalcin proposes. As before, we relativize interpretations to a world and information-state parameter.[21] The clauses for modals (which we presume only admit of epistemic interpretations) look familiar.

[21] Yalcin also relativizes extension assignments to a context parameter, but this won't matter for the cases we consider.

$$[\![\Box\phi]\!]^{w,i} = t \Leftrightarrow \forall w' \in i : [\![\phi]\!]^{w',i} = t$$

$$[\![\Diamond\phi]\!]^{w,i} = t \Leftrightarrow \exists w' \in i : [\![\phi]\!]^{w',i} = t$$

Yalcin pairs these with a familiar Hintikkan semantics for attitude reports that treats attitude verbs as quantifying over the worlds compatible with the truth-conditional information contained in an agent's attitude state. But he introduces a slight modification: in addition to having the attitude verb quantify over worlds compatible with an acceptance state, that verb also shifts the information-state parameter relative to which the evaluation takes place. Letting

$$S_x^w = \text{the set of worlds not excluded by what } x \text{ supposes in } w$$

we have

$$[\![x \text{ supposes } \phi]\!]^{w,i} = t \Leftrightarrow \forall w' \in S_x^w : [\![\phi]\!]^{w',S_x^w} = t.$$

For an information-state-insensitive complement ϕ expressing a set of truth conditions given by p, the agent will count as supposing the complement just in case the worlds compatible with their supposition state are all p-worlds. But with a modalized complement, things change. The modal effectively alters the quantificational force relative to which the modal prejacent (i.e. the ϕ in $\Diamond\phi$) is evaluated. For example,

$$\begin{aligned}[\![A \text{ supposes } \Diamond\phi]\!]^{w,i} = t &\Leftrightarrow \forall w' \in S_x^w : [\![\Diamond\phi]\!]^{w',S_x^w} = t \\ &\Leftrightarrow \forall w' \in S_x^w : \exists w'' \in : S_x^w [\![\phi]\!]^{w'',S_x^w} = t \\ &\Leftrightarrow \exists w'' \in S_x^w : [\![\phi]\!]^{w'',S_x^w} = t \vee S_x^w = \emptyset\end{aligned}$$

Accordingly, to request that someone suppose an epistemic contradiction is to request that they get into a contradictory state of mind. For to suppose a truth-conditional ϕ is to suppose in ways that rule out all $\neg\phi$ worlds. And to additionally suppose $\Diamond\neg\phi$ is to suppose in a way that either does not rule out all $\neg\phi$-worlds or supposes away all worlds. Since the state cannot do the first it

must do the second.[22,23]

Yalcin obtains a parallel result for conditional antecedents by adopting a semantics similar to that we saw K&M use. We define the following update on an information state by a sentence.

$$\text{Let } i +_y \phi = \textsc{max}\, i' \subseteq i : [i' \neq \emptyset \wedge \forall w' \in i' : [\![\phi]\!]^{w',i'} = t]$$

Note that Yalcin's update will not return a value when there is no unique maximal subset of i. Also, unlike with K&M, he builds in a provision that the subsets considered in an update must be non-empty. A conditional checks for truth of a consequent relative to the worlds of the shifted information state (evaluated relative to the shifted state as well).

$$[\![\phi \rightarrow \psi]\!]^{w,i} = t \Leftrightarrow \forall w' \in i +_y \phi : [\![\psi]\!]^{w',i+\phi}$$

This gives the result that conditionals embedding epistemic contradictions as antecedents are never true. This is because the update $i +_y [\phi \wedge \Diamond\neg\phi]$ is undefined, as no non-empty information state can satisfy the constraints imposed by the epistemic contradiction, for the same reason as we saw with supposition states.

Let me return to the treatment of attitude states. A fascinating thing about Yalcin's domain semantics is that there is no set of truth-conditions corresponding to a modalized claim like $\Diamond\phi$ (for non-trivial truth-conditional ϕ), such that when one believes $\Diamond\phi$ one's attitude state subsumes those truth-conditions. That is, we can show that there is no truth-conditional proposition p such that one accepts $\Diamond\phi$ just in case one's acceptance state rules out

[22] A *very* minor, and I hope irrelevant, difference between my exposition and Yalcin's is that Yalcin says "no state of supposition S" satisfies the ascription and asking someone to suppose an epistemic contradiction is to "request [they] enter into an impossible state of supposition, a request that cannot be satisfied" (Yalcin, 2011, 996). *Technically* this is only true if the null state is impossible: the initial quantification in the truth-conditions for the ascription are only vacuous if the state is non-null. In what follows, I will sometimes omit discussion of mental states with null truth-conditional content for ease of exposition.

[23] In Yalcin (2011), Yalcin refines this picture of attitude ascription by giving attitude states question-sensitive structure. To accept $\Diamond\phi$ (for truth-conditional ϕ) accordingly requires not only that one's acceptance state leave open ϕ-worlds, but also that one be 'sensitive to the question' of whether or not ϕ. While the sensitivity of an acceptance state to a question is an important matter that could even be tied to questions about reasoning broadly construed, I do not think it has direct relevance to *inference* as I have been investigating it in this book. Accordingly, in part to simplify exposition, I will ignore question-sensitivity. Many of the claims in what follows could be secured by restricting attention to the *omininquirent*, or all inquiring, agents.

every non-p-world. For either such a p would be true at all worlds, or not. If p were true at all worlds, every state should accept $\Diamond\phi$, which is not true (as any state with no ϕ-worlds fails to accept $\Diamond\phi$). And if p were false at some worlds, then the trivial belief state (which rules out no worlds) would not accept $\Diamond\phi$. But the trivial state *does* accept $\Diamond\phi$, in virtue of having some ϕ worlds.

Yalcin (2011) motivates his treatment of attitude ascriptions within a broader expressivist framework for epistemic modality. The view works roughly as follows: Sentences that vary in their evaluation exclusively along the information-state parameters (as pure modalized claims do) pick out a property of information states. The property is: being one of the information states at which the sentence evaluates to truth. Since acceptance states can be modeled as bodies of truth-conditional information, we can accordingly take the sentences in question to pick out a property of mental states. Then we can take an assertion of the sentence to constitute a recommendation to get into a mental state with that property. As we've seen, there need be no single proposition (or set of propositions) that one needs to accept in order for one's acceptance state to have the property in question. For example, an assertion of $\Diamond\phi$ would recommend that one get into a belief state that does not rule out ϕ. This is achieved not by believing some proposition, but by *not* believing $\neg\phi$. In this way we can make sense of the effects of asserting a modalized claim without assigning what is asserted a truth-conditional content, let alone a truth-value. Effectively the same story can be told for sentences that vary non-trivially in their evaluation both with respect to the information-state parameter and the world parameter.

As this review reveals, Yalcin's expressivism is predominantly a theory of the conventional assertoric effects of information-state-sensitive language. It is worth emphasizing that attitude reports, on Yalcin's semantics, are not information-state-sensitive in this way.[24] Because an attitude verb effectively binds the information state parameter, the evaluation of an attitude report is only sensitive to the world parameter. (Note, e.g., that in the elaboration of the semantics for "A supposes $\Diamond\phi$" given above, the information state parameter i figures nowhere in the right-hand-side conditions for its evaluation.)

The foregoing picture of mental states characterized by modalized lan-

[24] Or, at least, non-factive attitude reports are not so-sensitive. See Yalcin (2012b) for a discussion of the idea that knowledge attributions with probabilistic complements cannot express possible worlds truth-conditions.

guage contrasts with a picture on which believing a claim epistemically possible involves believing a fact about some body of information. A view of this contrasting kind might say, for example, that to believe ϕ is epistemically possible is to believe that one or more states of knowledge contain information compatible with ϕ. And that is simply to believe a truth-conditional proposition to the effect that the world is a particular way. On Yalcin's alternative:

> …there is no proposition that $\Diamond\phi$ at work. There are no '$\Diamond\phi$-worlds'. The question of whether A believes that $\Diamond\phi$ is just the question whether A's belief worlds leave open possibilities wherein the proposition that ϕ is true. To believe Bob might be in his office is simply to be in a doxastic state which fails to rule out the possibility that Bob is in his office. It is a first-order state of mind. …Such beliefs do not correspond to a distinctive class of believed contents; rather, they correspond to a distinctive way of being doxastically related to a proposition.
>
> (Yalcin, 2011, 309, footnote suppressed)

In this and similar passages Yalcin emphasizes that there is not a distinctively modalized (truth-conditional) content that one believes in order to have one's beliefs correctly described with a modalized expression. But there is a further point to be emphasized here that will be critical for understanding how to approach the issue of deductive inference in Yalcin's framework. For in that framework, apparently, *there is no further attitude state* beyond those states that relate to ordinary truth-conditional propositional content. One way to believe $\Diamond\phi$ for truth-conditional ϕ is to believe literally nothing at all. Another is to believe some batch of propositions, as long as they do not include $\neg\phi$. We can list the states that count as $\Diamond\phi$-states by listing only states bearing ordinary truth-conditional content. Once we have listed these, it appears we are done: there are not *further* states to list corresponding to modal acceptances. They have already been accounted for.

This point may be muddied a little by Yalcin's (tentative) framing of his view within a simple possible-worlds framework for modeling the content of attitude states. We assumed that an acceptance state like a supposition state could be modeled by a set of metaphysically possible worlds, which embodies the presumption that the holder of the state is logically omniscient: to suppose a truth-conditional proposition is immediately to suppose all of its truth-

conditional entailments. But it is worth noting that to preserve Yalcin's insights, we would need constitutive ties between first-order acceptance states and modalized acceptances, *even if* we refined attitude states so that they were borne to modalized contents.

This point is appreciated and discussed in MacFarlane (2014). MacFarlane sees in Yalcin's rejection of distinctively modalized contents obstacles in accounting for intuitively valid inference patterns like the following, which are often used as evidence for the need to posit shareable contentful objects of attitudes.

(10) (a) It might be raining.
(b) So, that it might be raining is true. [from (a)]
(c) Joe believes that it might be raining.
(d) So, Joe believes something true. [from (b,c)]

(d) appears to quantify over some truth-evaluable object—what would ordinary be taken to be a propositional content. But Yalcin's system does not provide for such objects of evaluation.

After setting up this issue, MacFarlane wonders what would become of Yalcin's view if it were modified to allow modalized attitudinal contents to account for argument forms like those in (10). He says:

> Yalcin's view [even if Yalcin were to embrace modalized contents] would retain its core expressivist commitment: its *identification* of believing that $\Diamond_e P$ with not believing $\neg P$, while being sensitive to the question of whether P.[25] On this kind of view, it is *conceptually impossible* to believe that it might be raining while believing that it is not raining, or to fail to believe that it is not raining (while being sensitive to the question) without believing that it might be raining.
>
> ...
>
> Relatedly, on Yalcin's view, any creature with the conceptual resources to believe that it is raining can also believe that it might be raining. For believing that it might be raining *just is* being sensitive to the question of whether it is raining and failing to believe that it is not.

[25] On this qualification concerning being sensitive to a question, see n.23 above.

(MacFarlane, 2014, 278-9, footnote suppressed)

I think MacFarlane is right about this. The constitutive ties are really inseparable from the work that Yalcin puts his framework for modals to. And this aspect of Yalcin's overarching view has critical implications for our understanding of how to model good inference in the presence of epistemic modal language. As noted in §II.1, Yalcin tentatively embraces an informational conception of entailment. He touts as a virtue of his semantics, on that conception of entailment, that we validate the following "intuitive pattern of inference."[26]

Łukasiewicz's principle[27]: $\neg\phi \models_i \neg\Diamond\phi$

Could Łukasiewicz's principle capture a good inference in *my* sense of inference? Curiously, it seems incapable of modeling a *good* inference in Yalcin's framework for being unable to model *any* inference type at all. To infer as I have understood it—and really on almost any reasonable understanding—requires passing from an acceptance of premises to acceptance of a conclusion. But on Yalcin's view as stated, this cannot happen in this instance. There is no *single* state of 'accepting $\neg\Diamond\phi$' which could stand as the conclusion of an inference of this kind. And even if there were, as MacFarlane notes, one would be in this state already, simply in virtue of accepting $\neg\phi$. To accept the latter *just is* to count as accepting $\neg\Diamond\phi$. To say one accepts the latter is just to describe aspects of the former acceptance. So there is nothing here to count as an inference, let alone a good one. There is no cognitive achievement in being in a state that you are already in.

It is worth noting that similar issues arise given Yalcin's semantics for conditionals. The following is informationally valid as we've already had occasion to note.

Modus Ponens? $\phi, \phi \to \psi \models_i \psi$

But is it possible to infer the consequent here from the premises, and on the basis of them? Just as with modals, truth-at-a-point for a conditional is only sensitive to variation in an information-state parameter. Accordingly, unless we give a special treatment to such expressions, they would fall under the purview of Yalcin's expressivism (whether modals are embedded in the conditional or not).

[26] Yalcin (2011, 1005).

[27] So-called because Łukasiewicz appears to endorse it in Łukasiewicz (1930/1970).

Focusing for now on truth-conditional ϕ and ψ, to accept $\phi \rightarrow \psi$ is to be in an attitude state that has only ψ-worlds among the ϕ-worlds it accepts.[28] Unlike with Łukasiewicz's principle, Modus Ponens? *can* have non-modalized conclusions which could in principle serve as the conclusion of an inference that one hadn't yet accepted, even if one accepts the premises.[29] The problem is now that the conditional does not represent a possible *premise-attitude* on which the conclusion could be based. It is merely a re-description of a state one is in, in virtue of holding attitudes to other contents. The resulting view thus avoids invalidating *inferential* Modus Ponens for the natural language conditional, but apparently at the cost of making it impossible to perform an inference worthy of that name.

It is worth stressing again that none of this is a result of Yalcin's (perhaps tentative or provisional) use of a possible-worlds framework for mental content, which familiarly presupposes some form of logical omniscience. The problems would persist even if we introduced distinctively modalized contents as possible objects of attitudes, and it would persist for the very reasons MacFarlane emphasized above. The problem is a result of something *foundational* for Yalcin's account of the kinds of mental states we are in when we are correctly described using modalized complements.

This still leaves us with a key question: if Yalcin's view were correct, what would a logic tracking conditions on good inference look like? If we provisionally ignore the logic of attitude reports themselves, the answer is simple: it is the logic that governs the fragment of the language that is modal- and conditional-free. Nothing, for example, prevents it from being a purely classical logic. What counts as a good inference cannot change merely from new ways of grouping old mental states into classes—however those classes are formed.[30] It is worth adding that it is not obvious that integrating Yalcin-style modals or conditionals alters our conception of good reasoning more broadly either. As MacFarlane notes, one doesn't even need modal concepts to accept modalized claims on Yalcin's view. What counts as good reflective reasoning in the modalized context should thus accessible to an agent without 'modal

[28] Perhaps additionally: the state must be sensitive to the question of whether ϕ, or whether ψ, or some composite of them—again see n.23.

[29] Or, at least, this is possible if we integrate the possibility of modalized contents into Yalcin's framework.

[30] Cf. the related lesson for logics in the presence of unresolved lexical ambiguity in §10.4, which we will return to again below.

concepts,' if it even makes sense to speak of those. And that seems to mean that the constraints on such reasoning should be statable without the help of modal locutions. Perhaps some of these rules could be *re*-described with the help of modal locutions (as we will see below). But this cannot generate new forms of reasoning. They would just be new ways of describing old forms of reasoning.

It is worth emphasizing two things about the foregoing conclusions. First, there is an important sense that the view I've been attributing to Yalcin is an unrefined version of his more considered views. To deal with probabilistic modals, Yalcin ultimately models acceptance states as credal states (and the more recent trend is to model such states with sets of credence functions). I'll come back to discuss what effects this kind of refinement has on the view in §11.5. (To preview: it makes matters much worse.)

Second, even on the 'unrefined' picture, we should bear in mind that there are other relations of theoretical interest besides those that *directly* undergird good deductive inference, some of which could well merit the title of 'logics.' We should not overstate the scope of the above lesson by ignoring them. So before turning to MacFarlane's views, let me say a little about some relations of this kind.

What other conceptions of logic could be applied within Yalcin's framework? As we saw in Chapter 10, Kaplan carved out space for a construal of logic that was independent of good inference: roughly, the investigation of truth-no-matter-what-context-a-sentence-is-used. But his definition appealed to a notion of truth-at-a-context which won't obviously be of use here. Yalcin's expressivism eschews the idea that context initializes a value for an information-state parameter. Since assertions of modal claims aren't evaluable for truth, it is not easy to extend the Kaplanian idea to this setting without taking up commitments Yalcin wants to avoid.

Instead of focusing on truth relative to a context, we could investigate relations of rational acceptance relative to a single agent's information state. Suppose I am in a given state of belief. We can ask: *given* what I currently accept, what is it rational to accept further? We could try to interpret Modus Ponens[?] as telling us that to the extent someone accepts the premises, it is rational for them to accept the conclusion. As we've seen in Chapter 3, this probably won't lead to useful results. This attempted formulation of a 'logic' will face all the typical problems for developing logico-normative bridge prin-

ciples for acceptance states on the basis of deontic language. (For example, will the Preface Paradox provide a counterexample to a generalized form of conjunction introduction on this approach?) As noted in that chapter, attempts to formalize relations like rational acceptance, in spite of increasingly baroque qualifications, continue to be subject to basic counterexamples. And the increasing qualifications tend to obscure the contribution from anything that could be considered a 'logic' and instead to incorporate more general epistemic norms, including those of prudence, responsiveness to evidence, and so on. So I'm not sure this is a fruitful path to pursue either.

In spite of these obstacles, there is at least one other construal of the consequence relation—or really a family of consequence relations—that I think could be fruitful to explore in the context of something like Yalcin's framework. To formulate them, it will help to add some refinement to the framework's use of sets of worlds to model acceptance states, since these model agents bearing those states as logically omniscient. On the refinement I have in mind, an acceptance state corresponds to a set of propositions—abstract 'correctness-evaluable' objects of attitudes. But I suspect the idea behind the ensuing construal of consequence could be developed in alternative ways (e.g., though the use of fragmentation).[31]

I will accordingly provisionally make use of the following adjustments and assumptions:

(I) Propositions are abstractions individuated at least as finely as sentences of our language. A sentence ϕ expresses a proposition $|\phi|$ which in turn determines a set of accuracy conditions $[\phi] = \{\langle w, i\rangle \mid [\![\phi]\!]^{w,i} = t\}$.

(II) A proposition $|\phi|$ is *truth-conditional* just in case

$$\forall w, i, i' : \langle w, i\rangle \in [\phi] \Leftrightarrow \langle w, i'\rangle \in [\phi]$$

$|\phi|$ is *information-sensitive* otherwise.

(III) A *narrow acceptance state* $\overline{A}$ is given by a set of truth-conditional propositions. A narrow acceptance state $\overline{A}$ determines a body of information compatible with it,

$$\overline{A}_I = \{w \mid \forall |\phi| \in \overline{A}, \exists i : \langle w, i\rangle \in [\phi]\}$$

[31] On fragmentation see e.g. Lewis (1982) and the discussion in Chapter 4.

(IV) The *informational closure* of a narrow acceptance state $\overline{A}$ is

$$C(\overline{A}) = \overline{A} \cup \{|\phi| \mid \phi \text{ is information-sensitive and } \overline{A}_I \rhd \phi\}$$

(V) An *acceptance state* A is the informational closure of some narrow acceptance state. The body of information determined by an acceptance state, A_I, is identical to that determined by the narrow acceptance that generates it.

The foregoing claims are meant to capture the following ideas. An agent's acceptances are 'fundamentally' given by a set of propositions whose accuracy (intuitively truth in this case) varies only from world to world. That is, the agent's acceptances are fundamentally given by what I call a 'narrow acceptance state.' But there are also further propositions whose accuracy varies with changes in an information-state parameter. Whether an agent accepts one of these propositions is assumed to be fully determined by their set of more fundamental truth-conditional acceptances, via relations of support. This is a plausible way of understanding how information-sensitive sentences continue to not really represent 'further things to accept,' even once we have adopted a framework accommodating finer propositional contents—as per MacFarlane's suggestion above. One only counts as believing these further propositions in virtue of the ordinary truth-conditional structure of one's acceptance state. This implementation does represent some substantive choices about how exactly to do this. For example, this proposal treats propositions expressed by sentences which are sensitive to *both* world and information-state parameters in the same way it treats those expressed by sentences which are *only* sensitive to an information-state parameter. One could explore a more nuanced treatment, but it will be unnecessary to explore the consequence relations of interest to me.

Against this backdrop we can consider the following relation among sentences.

ϕ is an open consequence of Γ iff either

(i) ϕ is information-sensitive, and for any acceptance state A: if $|\gamma| \in A$ for every $\gamma \in \Gamma$, then $|\phi| \in A$; or

(ii) ϕ is truth-conditional, and for any acceptance state A: if $|\gamma| \in A$ for every $\gamma \in \Gamma$, then there are truth-conditional propositions in

A that necessitate $|\phi|$.[32]

Open consequence is at two removes from the kind of entailment relations that my conception of logic is concerned with. First, open consequence is disjunctive: it tracks *either* relations of constitution *or* necessitation relations supporting the possibility of good inference. Second, even when it tracks relations supporting good inference it does so in a *schematic* way: the inferential relations tracked may be undergirded by contents that do not correspond directly to the sentences related by open consequence. In fact, a single set of sentences related by open consequence may correspond to an infinite set of significantly distinct inferential transitions.

To appreciate the first point, we can note that instances of Łukasiewicz's principle are related by open consequence in virtue of relations of constitution. If $|\neg\phi|$ is in an acceptance state then, *ipso facto*, so is $|\neg\Diamond\phi|$. To accept $|\neg\phi|$ just is one way of accepting $|\neg\Diamond\phi|$. No inference between them is necessary. In fact, none is possible.

To appreciate the second point, we can note that instances of Modus Ponens$^?$ are also related by open consequence. Suppose ϕ and ψ express truth-conditional contents. Then there is no single set of truth-conditional contents in virtue of which an acceptance state would contain $|\phi \to \psi|$. But still, as long as one accepts $|\phi|$, then no matter how one counts as accepting $|\phi \to \psi|$ there will always be some possible good deductive inference from one's truth-conditional acceptances to $|\psi|$. For example, one could believe $|\phi \to \psi|$ by believing $|\neg\phi \vee \psi|$. From that latter premise and $|\phi|$ one can infer $|\psi|$ by a good inference—namely disjunctive syllogism. Or one could also believe $|\phi \to \psi|$ simply by believing $|\psi|$. Though it is trivial, there is a good inference from the premises $|\phi|$ and $|\psi|$ to $|\psi|$. But the inference pattern in this case is of course no longer an instance of disjunctive syllogism. This is what I meant by saying that open consequence is a *schematic* deductive relation: Modus Ponens$^?$ actually represents a *class* of importantly distinct good inferential transitions.

This second qualification on open consequence comes with a corresponding lesson about the utility of open consequence in the study of deductive inference, familiar from Chapter 10. There, in discussing logics for unresolved

[32] I.e. any world paired with some information-state among the accuracy conditions of all the truth-conditional propositions in question is also paired with some information-state among the accuracy conditions of the conclusion ϕ.

lexical ambiguity and some forms of context-sensitivity, we saw that a logic could investigate coarser and coarser classes of relations undergirding good deductive inference only at the expense of losing track of *possible bases* for such inference. The same problem arises here. When one agent infers $|\psi|$ from $|\phi|$ and $|\psi|$, and another agent infers $|\psi|$ from $|\phi|$ and $|\neg\phi \vee \psi|$, what unites these good inferences together *beyond* that they both necessarily preserve truth? It seems like there is nothing else interesting in common to these inferences—no common 'mode of inference' or common basis for inference. The grouping of these inferences together by open consequence is, accordingly, somewhat artificial. So not only does MODUS PONENS$^?$ fail to 'directly' model any possible inference type, it also cannot helpfully indirectly classify inferences into a unified type on grounds of common inferential bases. In line with my ecumenical stance on broader uses of the term "logic", I have no objection to the investigation of relations of open consequence, and one can even call the result a 'logic' if one likes. The important point is to be clear about what one is doing by investigating such classifications, and about what the limitations of the classifications are.

Suppose one were interested in relations of open consequence. How would they be formalized? Since ŁUKASIEWICZ'S PRINCIPLE and MODUS PONENS$^?$ satisfy open consequence, it is natural to ask if open consequence is equivalent to (or 'modeled by') informational consequence. In the framework just given this would at least hold provided that for every information state i there is some truth-conditional proposition $|\phi_i|$ with i as its truth-conditions.

> Suppose that every information state i there is some truth-conditional proposition $|\phi_i|$ with i as its truth-conditions. Then:
>
> $$\Gamma \models_i \phi \Leftrightarrow \phi \text{ is an open consequence of } \Gamma$$
>
> $\Rightarrow$: Given an acceptance state $A = C(\overline{A})$, suppose $|\gamma| \in A$ for every $\gamma \in \Gamma$. Then $\overline{A}_I \rhd \gamma$ for every $|\gamma| \in A$. Since $\Gamma \models_i \phi$, $\overline{A}_I \rhd \phi$. Then if $|\phi|$ is information-sensitive, $|\phi| \in C(\overline{A}) = A$ by definition. And if $|\phi|$ is truth-conditional, $\overline{A}_I \subseteq \{w \mid \exists i : [\![\phi]\!]^{w,i} = t\}$. This can only occur if $\overline{A}$ (and so A) contains truth-conditional propositions necessitating $|\phi|$.
>
> $\Leftarrow$: Consider some information state i such that $i \rhd \gamma$ for every $\gamma \in \Gamma$. Let $|\phi_i|$ be a truth-conditional proposition with i as its truth-

> conditions. Let $\overline{A}$ be the narrow acceptance state consisting of $|\phi_i|$ and every truth-conditional $|\gamma|$ such that $\gamma \in \Gamma$. Then by design $A = C(\overline{A})$ is such that $A_I = \overline{A}_I = i$, and $|\gamma| \in A$ for every $\gamma \in \Gamma$. If $|\phi|$ is information-sensitive, the assumption of open consequence assures $|\phi| \in A$. In this case $A_I \rhd \phi$, i.e. $i \rhd \phi$. If $|\phi|$ is truth-conditional, the assumption of open consequence assures that there are truth-conditional propositions in A that necessitate $|\phi|$. But i is a subset of the intersection of the truth-conditions of those necessitating truth-conditional propositions, and as such $i \rhd \phi$.

What this shows us is that while informational consequence cannot 'directly' model good inferential relations, it could (on certain refinements of Yalcin's views) be used to indirectly model relations undergirding good inference alongside relations of constitution in the somewhat limited way that open consequence does.

It is important to know that when we interpret informational consequence as a means of tracking open consequence it no longer stands as a 'rival' logic to (say) classical logic or a broadly Kaplanian logic as regards the tracking of inference. It is instead a *supplementary* and *compatible* mode of tracking inference at a higher level of abstraction (analogous, e.g., to the assignment of logical properties to logical schemas), afforded now by the modal and conditional devices that are used to classify acceptance states. There is no harm in simultaneously adopting both logics, as long as one is clear about the linguistic fragments to which they apply, and the different ways they model features of mental content.

Before leaving off Yalcin I want to mention one quick additional point. First, I said above that if we "ignore the logic of attitude reports themselves," then the most perspicuous logic for Yalcin's framework is the one that "governs the fragment of the language that is modal- and conditional-free." The caveat there is important, as Yalcin's semantics for attitudes, coupled with that for modals and conditionals, can lead to some unusual entailments among sentences containing attitude verbs that wouldn't appear in the absence modalized and conditionalized attitudinal complements. These entailments have been used to raise questions about Yalcin's framework by Schroeder (2015, Ch.9). In particular, Schroeder notes that Yalcin's framework seems to validate certain 'importation' and 'exportation' principles for connectives and attitude ascrip-

tions. For example, Yalcin's view seems to entail that for an agent A who is reflective and, rational[33] belief in a disjunction with a modalized complement $\textsc{mod}\psi$ will entrain belief one of the disjuncts:

$$\forall i, w\text{: if } [\![Bel_A(\phi \vee \textsc{mod}\psi)]\!]^{w,i} = t, [\![Bel_A(\phi) \vee Bel_A(\textsc{mod}\psi)]\!]^{w,i} = t$$

I think it is natural to phrase the relation between attitude states uncovered here in terms of entailment, including logical entailment (if attitude verbs can count as logical vocabulary). Unlike with unembedded modals, there is no reason to avoid this characterization. After all, non-factive attitude ascriptions on Yalcin's framework have ordinary truth-conditions—i.e. are information-state insensitive—because the attitude verb effectively binds the information-state parameter. So it makes sense to evaluate these commitments of Yalcin's view in terms of what one can or cannot deduce, inferentially, from the fact that a suitably reflective agent believes a disjunction. In this respect the 'full' logic for deduction for Yalcin's framework might not consist of the modal- and conditional-free component of his language, but the component which is modal- and conditional-free only outside of attitude contexts. There is a distinctive logic for attitudes here, and one that is quite controversial.

There is of course much more to say here about how to develop logics within something like Yalcin's framework for modals, conditionals, and attitudes—I have barely scratched the surface here. But more than these possible details, I am interested in the *contrasts* between the forms of logic that we encounter on Yalcin's view and those we encounter within rival frameworks for information-state semantics. So let's see how questions about inference take a wholly different shape within a different approach to modals and conditionals given by MacFarlane.

11.2.2 MacFarlane's Assessment Sensitivity

MacFarlane (2014) explores the prospects for a relativist treatment of epistemic and deontic modals and conditionals. On this view, uses of such expressions are *assessment-sensitive*, which is to be understood by analogy with more familiar context-sensitivity or what MacFarlane calls *use*-sensitivity. Just as a context of use is a possible circumstance in which a sentence could be used, a context of assessment is a circumstance relative to which a sentence can be

[33] And so is suitably sensitive to relevant questions—see n.23.

'assessed' by an evaluator. The idea is that contexts of assessment can play a role just like contexts of use in our compositional and post-compositional semantics. For example, just as some expressions may have their extensions relativized to a context of use, others may have their extension relativized to a context of assessment. And just as context can play a role in a definition of truth-of-a-sentence-at-a-context-of-use (as we saw Kaplan give in Chapter 10), so too a context of evaluation can enter into a definition truth-of-a-sentence-at-a-context-of-use-and-context-of-evaluation.

While there is much background that goes into clarifying this kind of relativism and its applications, it is best here to simply delve in and see how the framework is built up for modals. MacFarlane's compositional semantics recursively defines a notion of truth at a point of evaluation consisting of a context (of use), a world, a time, an information-state, and a variable assignment. Since context of use, time, and variable assignment will be irrelevant to the language fragments of interest to us, I will suppress these parameters. Then the compositional theory for epistemic modals will assign them semantic values that should now look increasingly familiar.

$$[\![\Diamond_e \phi]\!]^{w,i} = t \Leftrightarrow \exists w' \in i : [\![\phi]\!]^{w,i} = t.$$

$$[\![\Box_e \phi]\!]^{w,i} = t \Leftrightarrow \forall w' \in i : [\![\phi]\!]^{w,i} = t.$$

The semantics for MacFarlane's conditional also looks similar to recently seen proposals. We can define an informational update for MacFarlane that looks like Yalcin's, except that instead of producing a unique maximal state updated with a conditional antecedent, it produces a set of such states. The conditional then simply checks whether a consequent supports all such maximal states.

$$\text{Let } i +_m \phi = \{i' \subseteq i \mid i' \rhd \phi \text{ and } \neg\exists i'' \supset i' : i'' \rhd \phi\}$$

$$[\![[\text{if } \phi]\psi]\!]^{w,i} = t \Leftrightarrow \forall i' \in i +_m \phi : i' \rhd \psi$$

Just as with K&M's conditional, this conditional checks a consequent against maximal antecedent-supporting information states. Unlike that conditional, and like Yalcin's, it checks for support of the consequent (rather than checking for its truth relative to the information state at the original world of evaluation).

None of these clauses, of course, does anything to distinguish the framework here as relativist. For example, the clauses for modals are just the clauses

we saw Yalcin appeal to in his expressivist framework. And were there an 'information-state-of-the-context-of-use,' these clauses could equally be exploited by a non-relativist form of contextualist descriptivism.

Relativism enters the picture in what MacFarlane calls the *postsemantics*, which allows contexts of assessment to initialize values for the information-state parameter when defining truth for a sentence. The relativist has several options here, but a simple one will help illustrate the idea.

> Solipsistic Relativist Postsemantics. A sentence S is true as used within a world w and assessed from context c just in case
>
> $$[\![S]\!]^{w,i_c} = t$$
>
> where i_c is the information state determined by what is known by the agent of at c at the time of c.

Finally, we can tie the post-semantics to a pragmatics for assertion and retraction as follows.

> Reflexive Truth Rule. An agent is permitted to assert that p at context c_1 only if p is true as used at c_1 and assessed from c_1.
>
> Retraction Rule. An agent in context c_2 is required to retract an (unretracted) assertion of p made at c_1 if p is not true as used at c_1 and assessed from c_2.

To see how the framework might be applied, consider a simple case like the following.[34]

> Coffee Shop
> *Sally*: Joe might be in China. I didn't see him today.
> *George*: No, he can't be in China. He doesn't have his visa yet.
> *Sally*: Oh, really? Then I guess I was wrong.

Dialogs like this have seemed puzzling to some. Sally can seem warranted in making her assertion. This would be true if her assertion were just about the knowledge she had. But then why would she retract her earlier claim upon learning more? E.g., if Sally had led by saying "for all I know, Joe might be in

[34] Drawn from MacFarlane (2014, 240).

China", then it would be bizarre for her to take that claim back upon learning more, since it was still true, at the time of asserting, that for all she knew Joe was gone.

The relativist can make sense of both Sally's warrant and her tendency to retract. She is warranted in making her initial assertion since it is true as assessed from the context of assessment in which she speaks, precisely because at that time Sally did not know Joe wasn't in China. So by the Reflexive Truth Rule Sally is permitted to make her assertion.

In spite of this, Sally could still be obligated to retract her earlier assertion when she learns more from George. This is because by the time Sally has heard from George, she is now in a context of assessment relative to which the earlier assertion is false, because at that time Sally (who is still speaker) now knows Joe is not in China. So relative to that information present in the new context of assessment, the *original* modalized claim is false. Retraction Rule then tells us that Sally is required to retract her assertion as a result.

Now there is a *lot* more to say about cases like Coffee Shop, both for and against relativist positions. And bear in mind that Solipsistic Relativist Postsemantics is hardly the only postsemantics the relativist can employ. I have no interest here in exploring the plausibility of relativism per se. Instead I want to know: if such a semantics and its applications pan out, what would that teach us about good inference, and a logic designed to track it?

As always, to investigate inference we must get clearer on what relativism teaches us about the mental states that inference would mediate between. For MacFarlane, sentences whose evaluation is sensitive to contexts-of-assessment (either in the compositional semantics, or via the postsemantics) can correspond to abstract propositional objects which share this sensitivity. For MacFarlane, we first make sense of genuine relative truth by examining linguistic use and norms for assertion and retraction. And once we do this, there are no obstacles to taking propositions—construed as what is asserted and what is believed—in relativistic terms.[35] Such propositions vary in their truth not only relative to a possible world, but also relative to a further parameter (in the case of interest to us, an information state) which can be filled in by a context of evaluation.[36] These "assessment-sensitive propositions can be believed,

[35] See, e.g., MacFarlane (2014, 114).

[36] MacFarlane is careful to note that the mere accommodation of such propositions is not only insufficient, but not necessary to be a relativist. Since he treats modalized propositions in relativist terms, though, these subtleties will not matter to me here.

judged, doubted, supposed, and so on" in just the way that ordinary propositions may.[37]

While MacFarlane is neutral on many features of propositions (e.g., whether they are linguistically structured), he does adopt a semantics for attitudes on which they express relations between agents and propositions.[38] Where $|\phi|$ denotes the proposition expressed by ϕ, we have.

$$[\![A \text{ believes that } \phi]\!]^{w,i} = t \text{ iff } A \text{ has a belief with content } |\phi| \text{ in } w.$$

This leads to an important contrast with Yalcin's approach. Recall that Yalcin's view encapsulates the idea that modalized attitude verb complements characterize classes of attitude states whose informational structure can be described without the help of modal locutions. This results in relations of constitution that hold between underlying acceptances (or the absence of various acceptances) of ordinary truth-conditional contents, and true descriptions of one's acceptance state using modals. MacFarlane is suspicious of this kind of framework, and highlights the relational semantics as enabling us to avoid its pitfalls. Continuing a quotation from §11.2.1:

> On [Yalcin's] view, it is *conceptually impossible* to believe that it might be raining while believing that it is not raining, or to fail to believe that it is not raining (while being sensitive to the question) without believing that it might be raining. For the relativist, by contrast, these are distinct states, and it is possible in principle to be in one without being in the other. Of course, on the relativist view one *ought not* be in one without being in the other. Given that one aims at believing what is true given one's evidence, and given the intension of *it might be raining*, it would be a mistake to believe this proposition while believing that it is not raining, and a mistake to fail to believe this proposition while considering whether it is raining and not believing that it is isn't. A mistake—but one it is possible to make, at least in principle.
>
> ...
>
> These differences are substantive. While there is something attractive about identifying believing that $\Diamond_e\phi$ with leaving-open that ϕ, such a view rules out a kind of "epistemic akrasia"

[37] MacFarlane (2014, 117).
[38] MacFarlane (2014, 156).

> that seems genuinely possible—the state of believing that $\Diamond_e\phi$ while simultaneously believing $\neg\phi$, and hence not leaving-open that ϕ. The relativist view leaves such combinations possible, while explaining what is wrong with them, and why they are rare.
>
> (MacFarlane, 2014, 278–9)

MacFarlane emphasizes the intuitive idea that there are rational, non-necessary connections reported between the sentences related in Łukasiewicz's principle. So unlike Yalcin, he takes belief reports with these sentences as complements to report *independent* attitude states that bear rational, and not constitutive, connections. If true, this is extremely important for the investigation of reasoning and inference: the possibility of *transitioning* between attitude states corresponding to the sentences of Łukasiewicz's principle appears to have been restored.

But matters here are far more complex than they may initially appear. In fact, by separating out elements that get run together in Yalcin's framework, a whole host of questions have arisen about their relationships. I do not think MacFarlane addresses all these questions, and they will turn out to matter for whether, and how, we can integrate an understanding of deductive inference into the relativist's picture of mentality.

To begin, note that there are now *three* separate features of an acceptance state S (of a given agent A) that are worth keeping track of.

> S_p: The set of propositions $|\phi|$ to which an attitude state S is related. (That is, the $|\phi|$ such that "A Vs that ϕ" is true, where "A" denotes the bearer of S, and "V" expresses the attitudinal S-relation.)
>
> S_i: The information contained in an acceptance state S, modeled by a set of possible worlds.
>
> S_α: The set of sentences α that are supported by the information contained in S (i.e. those ϕ such that $S_i \rhd \phi$).

Recall how these were related on the extension of Yalcin's view I offered in §II.2.1. Fundamentally, an attitude state is characterized by the truth-conditional propositions in S_p. From this, facts about S_i (and so S_α) are deriv-

able in straightforward ways:

$$S_i = \{w \mid \forall |\phi| \in S_p, \exists i : \langle w, i\rangle \in [\phi]\}; \text{ and}$$
$$S_\alpha = \{\phi \mid S_i \triangleright \phi\}$$

Before we took truth-conditional propositions to be a fundamental characterizer of mental structure. But now that there is freedom to believe an informational proposition without this *following* from one's truth-conditional beliefs, there is space created between S_p and S_i that makes it surprisingly hard to say what the relationship should be between them. Recall S_i is, by assumption, a set of worlds. Which? Could it be the information state that supports each ϕ such that $|\phi| \in S_p$? The problem is that there is no such unique state in general. If my sole belief is $\Diamond_e\phi$ for truth-conditional ϕ, then any state with a ϕ-world will do. Could we pick the smallest state supporting ϕ for each $|\phi| \in S_p$? There will always be a unique such state: the empty set. The empty set supports every sentence trivially. But obviously it would be a mistake to associate every mental state with contradictory information. Even if we ignore the null set there will be too many sets to consider. For example if my sole belief is $\Diamond_e\phi$, then any singleton ϕ-world will support my beliefs.

Perhaps instead of the smallest state accepting each ϕ such that $|\phi| \in S_p$ we should take the largest. But again there is not any unique such largest set in general. Recall $K\&M$'s example of "we ought to block a single shaft" in the context of the miner puzzle: this should be supported by an information state with only miners-in-shaft-A worlds, and also an information state with only miners-in-shaft-in B worlds. So there are two incompatible maximal sets to consider that could correspond to belief in the "we ought to block a single shaft" alongside other beliefs characterizing the scenario of the puzzle. Though MacFarlane (2014, Ch.11) refines the semantics of deontic modals from K&M that we saw in §11.1, it will continue to create sentences accepted by multiple, incompatible, otherwise maximal information states. Indeed, the same problem can already be seen for MacFarlane's treatment of epistemic modals. If my only belief is $\Box_e\phi \vee \Box_e\neg\phi$ for truth-conditional ϕ, this is accepted by a state with all and only the ϕ-worlds, and also by a state with all and only the $\neg\phi$ worlds. Neither of these is 'larger' than the other in the relevant sense.

We could try to cobble together the largest information states accepting

each ϕ such that $|\phi| \in S_p$. But how? The intersection will do no good: this will consistently generate the inconsistent null information state whenever someone accepts a sentence supported by several incompatible information states. Could we take the union of these states? This option is certainly more conservative. But for that reason it seems to render attitudes reported with informational language conspicuously inert.

Consider again the miner puzzle. Suppose I know the basic facts of the case, but I am a poor reasoner and I haven't yet settled that I should not block either shaft. Now (perhaps irrationally) I come to the belief that I ought to block shaft A or that I ought to block shaft B. This seems like a major change in my belief state. For example, though I don't know which of the two shafts to block, it might be worth going to fetch the sandbags to be ready. But on the current proposal, the information in my belief state hasn't changed at all. The information 'added' to my belief state is just the union of the miners-in-shaft-A worlds and the miners-in-shaft-B worlds. That is, of course, just all the worlds there were to begin with.

In connection with this, taking the union of maximal acceptances gives us a story about how S_p and S_i are related, but at the cost of severing a natural relationship between S_p and S_α—something that should lead us to question the position's coherence. MacFarlane seems to allow that $S_\alpha \not\subseteq \{\phi \mid |\phi| \in S_p\}$ and for legitimate reasons. Just because the truth-conditional information in one's belief state supports something, does not mean one has yet explicitly arrived at a belief in what is supported. We need something like this distance to make sense of the 'akratic' states MacFarlane wants to leave open. But on the proposal we are exploring, we could also have $\{\phi \mid |\phi| \in S_p\} \not\subseteq S_\alpha$—i.e. one could believe $|\phi|$ without one's belief state supporting ϕ—which seems much more problematic.

To see this, suppose I have a single belief: $|\Box\phi \vee \Box\neg\phi|$ for some truth-conditional ϕ such that ϕ and its negation are both possible, and where the relevant modality is epistemic. On the current proposal, coming to this belief has no effect whatsoever on the information contained in my belief state, B_i. Accordingly, B_i will not support $\Box\phi \vee \Box\neg\phi$. After all, this would require B_i to contain only ϕ worlds or only $\neg\phi$ worlds. But since I believe nothing aside from $|\Box\phi \vee \Box\neg\phi|$, B_i contains both kinds of worlds.

This might seem to be a minor technical bug. But it is worth noting that it would normally lead to further, seemingly more serious aberrations through

its interaction with a relativist postsemantics and pragmatics. Suppose, for example, that we are working within our SOLIPSISTIC RELATIVIST POSTSEMANTICS. Suppose further that I come to know 'merely' $|\Box_d$block-A $\vee$ $\Box_d$ block-B|, the proposition expressed by "I ought to block a single shaft", in the context of the miner puzzle (though any other sentence accepted by incompatible information states will do). Now, the REFLEXIVE TRUTH RULE tells me I am permitted to assert the sentence expressing the proposition I know only if the proposition is true as used and assessed from my context. But our grip on the proposition's being true comes from our grip on the sentence being true. And to check whether the sentence is true as used and assessed from my context, by the SOLIPSISTIC RELATIVIST POSTSEMANTICS, I check whether my compositional semantics for the sentence assigns it truth relative to 'the information state determined by what is known by the agent of at c'—that is, the information state given by my knowledge. But we have just seen that the sentence expressing what I know is *not* true relative to that information state on the current proposal for its structure. It is not clear what, besides the REFLEXIVE TRUTH RULE, could license the assertion. So I seem unable to assert the sentence, even though it expresses a proposition that by hypothesis I *know*. It is worth emphasizing that the problems here are not unique to the SOLIPSISTIC RELATIVIST POSTSEMANTICS. The problem will arise for some agent whenever the information targeted in the postsemantics *can* subsume the information in one or more mental states.[39]

So, given an attitude state S, what *is* S_i? MacFarlane does not tell us—he simply assumes there is some body of information corresponding to it. I've tried to flag that once we go in for a relationalist semantics for attitude verbs, accommodate the existence relativistic propositions, and further shirk Yalcin-style constitutivism, the assumption here is a non-trivial one.

Now, there are ever so many adjustments to the framework that we could pursue to avoid encountering the problems we have been. Not to mention that we could simply bite the bullet with respect to one of the options I've already reviewed. I do not want to delve deeper into these issues here. Instead, I want to step back and review how it was that we encountered the obstacles we are

[39] Perhaps we can try to get out of this particular problem by saying that one cannot *know* merely a single proposition like $|\Box_d$block-A $\vee$ $\Box_d$ block-B| (e.g., by testimony) in the miner case. But this doesn't seem to make much sense if knowing that proposition is informationally vacuous—i.e., if a mental state that comes to accept it has ruled out no new worlds, as we are assuming.

finding, and note some implications it has for an information-state logic for deduction.

What was the point of considering information states, and modeling them with sets of worlds, to begin with? The following motivating story appears to be lurking in the background. A first idea is that there are bodies of information in the world given by things like newspapers, maps, books, speeches, and even attitude states. And a second idea is that a class of modals, and perhaps conditionals as well, function to characterize or describe the structure of these bodies of information. A modal like $\Diamond_e\phi$ intuitively captures some sense in which a body of information is *compatible* with ϕ. $\Box_e\phi$ intuitively says what *follows* from it. Likewise, a conditional might seem to report what follows from a body of information that is provisionally augmented with new information.

The key questions are: Compatible *how*? Follows *how*? If the newspaper reports that "tomorrow it will be cloudy in the morning and it will rain in the afternoon", intuitively it is not 'compatible' with the information in the newspaper that there are no clouds in the morning. Intuitively it 'follows' from the information in the newspaper that it will rain in the afternoon. But the newspaper contains neither the single sentence "it will be cloudy tomorrow" nor "it will rain in the afternoon", nor does it directly expresses either relevant proposition. Instead, it contains a conjunction that expresses a conjunctive proposition. This is just a reminder that the intuitive notions of 'compatibility' and 'following' would be *broadly logical* ones.[40] They are concerned with what one can safely conclude, or not, given the ways various information is encoded in the relevant body of information.

This is why it makes perfect sense to model an information state with a set of worlds, rather than (say) a set of propositions—*even* if the body of information characterized by a modal expression can be articulated in propositional form (as MacFarlane allows). For we would only be interested in the propositions insofar as they were closed under some kind of entailment relation. The standard problem for using sets of worlds to model acceptance states—namely, that such states are not closed under entailment—is actually transformed into a virtue in this context. We want the semantics for our information-state-sensitive expressions to be responsive to information that is rendered 'logically omniscient', closed under consequence, etc. A set of worlds represents that kind of information most simply and perspicuously.

[40] Most likely they will be general notions of entailment, subsuming logical relations.

Or, at least, it does this *if all the information is truth-conditional*. This is why Yalcin can get away without encountering any of the troubles we find for MacFarlane. His broadly constitutivist treatment of attitudes characterized by information-state-sensitive sentences allows us to think that, fundamentally, there is nothing to an information state beyond the truth-conditional information it subsumes. In this context, entailment is just necessary preservation of truth, compatibility just non-entailment of a negation.

But MacFarlane explicitly eschews this constitutivism. Information-state-sensitive sentences represent *further* things to report or believe, beyond truth-conditional propositions. Accordingly, it stands to reason that they could create further effects—even indirectly—by being part of a body of information characterized by an information-state-sensitive expression. This leads to the question: When we have a belief state (or knowledge state, or book, etc.) whose information is articulated propositionally, how do we get from these *expanded* batches of propositional information to an 'information state' of the sort that we need to evaluate the use of information-state-sensitive expressions? As I say, this is ultimately a question about entailment: what follows from a given set of propositions, including information-sensitive ones?

We can see that we are now stuck in a tight circle. We have a compositional semantics (and even a postsemantics, and pragmatics) which should help us gain insight into entailment relations. But the problem is that the framework cannot be applied broadly until we know what the bodies of information that the information-sensitive expressions target are like. And to figure that out, we appear to need some notion of entailment. So we cannot get from our semantics directly to an entailment relation: we must *already* have an entailment relation on hand to apply the semantic framework in the first place.

As alluded to before, there may be many competing ways to break out of the circle, some more plausible than others. The important take-away message for now is that MacFarlane does not give us enough information to make headway. In other words, right now, questions about the 'correct' deductive logic for a framework like MacFarlane's do not admit of any definite answers.

Now that we have at least seen the rudiments of two competing information-state semantics for modal and conditionals, I want to step back and draw some general morals for approaching questions about logic in that setting.

Though I've barely touched upon the rich semantic frameworks devel-

oped by Yalcin and MacFarlane, I think even these brief investigations lend support to my main contention about information-state semantics in this section. This is that the question "what is the proper *deductive* logic for a semantics of this kind?" *has no sense* independently of a detailed framework that draws proper connections between such a semantics and the contents of attitude states that could figure in deductive inferential relations. There are several surprisingly different ways to draw those ties, even for a single compositional semantic framework for information-state-sensitive language.

As we've seen, the compositional semantics for modals in Yalcin and MacFarlane's frameworks is essentially the same. But Yalcin's framework seems to *preclude* modal and conditional discourse from reporting individual attitudes that could figure in deduction. I argued that from this perspective, if we bracket the logic of attitude verbs, deductive logic remains unaltered by the addition of modals and conditionals, so the most perspicuous representation of deductive relations should simply exclude this language. We can integrate modal and conditional language in a *kind* of deductive logic modeled by informational consequence. But this only seemed possible by reconstruing consequence as what I called "open consequence"—a high-level re-description of classes of inferences and constitutive relations between attitudes, which is compatible with the more perspicuous underlying logic that excludes modal and conditional language.

By contrast MacFarlane's framework carves out a place for individual, distinctively 'modalized' or 'conditionalized' contents that seem like they could figure as the starting- and end-points of inference. But exactly what those relations are depends integrally on the information contained in an attitude state which takes these contents as objects. As I've argued, I do not think MacFarlane gives us enough details to specify that information. So while his framework makes room for a distinctive 'information-state logic of deduction,' we still do not have enough materials to *extract* such a logic from the framework without making further controversial commitments. Indeed, any general application of MacFarlane's machinery in fact seems to presuppose we already have answers to questions about deductive logic for informational language already in place. So even within a framework like MacFarlane's not enough has been said to give logical questions a determinate answer.

And there are frameworks besides these. For example, there is room for an information-state semantics that treats modal and conditional discourse as

stating simple, unrelativized truth-conditional content, whose expression is facilitated by the initializing of the information-state parameter by a privileged body of information picked out in a context of utterance. Obviously, a deductive logic for this framework may look quite different from that for both Yalcin and MacFarlane.

If we want to know "what is the correct (inferential) logic for a given information-state semantics?" the answer can only be that there *is no* framework-independent answer to this question. Only in the context of a framework that draws explicit ties between the semantics and the structure of mentality, and describes how that structure shifts (if it even can) in response to the incorporation of information-sensitive language, can questions about the relationship between information-state-sensitive language and deductive inference be given a sense. And only once we have a grip on the sense of those questions can further queries about deductive logic gain any traction.

11.3 Informational Consequence and the Preservation of Truth

The extreme sensitivity of information-state logics of deduction to a broader embedding framework can make investigating such logics an extremely subtle matter. This is an increasingly important point, as recent times have seen an out outpouring of work that purports to defend particular logics for information-state semantics, attack their general applicability, or point to places for possible refinement. I have in mind work like Schulz (2010), Bledin (2014), Mandelkern (2020), and Santorio (2022). I do not think one finds in the work of these authors much of a sense that a broader embedding framework makes a difference to the points about logic they defend.

To be fair, I think that for many (perhaps even most) of the arguments made in the citations just given, mild caveats can be introduced, or even claims can be recast to not concern 'logic' (or at least a logic of deduction), with relatively little loss of force. For example, Mandelkern presents apparent McGee-style counterexamples to Modus Ponens applied to counterfactual conditionals—cases that would be of significance regardless of how we understand what logical consequence *is*. Still, sometimes the points made in this work on logics for information-state semantics appear to turn more substantially on the subtle framework-relative matters that cannot be so easily turned

aside.

Here, I want to apply pressure to a recent tendency to conflate two issues: (a) the rejection of logical consequence relations for information-state semantics that are *formalized* in terms of preservation of something like truth at a point of evaluation, and (b) the rejection of *intuitive* construals of validity and consequence as concerned with relations of (necessary) truth-preservation. Language indicative of this conflation is pervasive in the emerging literature and is, I will argue, quite misleading. Here, I will focus on Justin Bledin's work on informational consequence to illustrate what I think the problems are, as Bledin is admirably clear and explicit about these issues.

Bledin contrasts several formal consequence relations for an information-state semantics broadly similar to those above, among which two familiar relations stand out: what in §11.2.1 I called, following Yalcin, diagonal consequence, $\models_{Tr}$, and informational consequence, $\models_{I}$. (Bledin subscripts the first consequence relation with 'Tr' instead of 'd' and the second with "I", a usage I will follow in this section.)[41]

> $\Gamma \models_{Tr} \phi$ iff for every information state i and world $w \in i$:
> if $\forall \gamma \in \Gamma, [\![\gamma]\!]^{w,i} = t$, then $[\![\phi]\!]^{w,i} = t$

> $\Gamma \models_{I} \phi$ iff for each information state i:
> if $\forall \gamma \in \Gamma, i \rhd \gamma$, then $i \rhd \phi$.

Bledin's main criticism of the first relation is that it "invalidates some good deductive arguments,"[42] giving some now familiar examples like the following.

$\neg\phi \not\models_{Tr} \neg\Diamond\phi$

(P1) Professor Plum didn't do it.

(C) It's not the case that Professor Plum might have done it.

$\phi \vee \psi, \neg\psi \not\models_{Tr} \Box\phi$

(P1) Either Mrs White did it or Miss Scarlett did it.

(P2) Miss Scarlett didn't do it

(C) Mrs White must have done it.

[41] To make these properly *logical* consequence relations, Bledin actually relativizes interpretation to a model. Since this relativization doesn't matter for my points below I omit the parameter.

[42] Bledin (2014, 287).

$$\phi \to (\neg\psi \to \xi), \phi \not\models_{Tr} \neg\psi \to \xi$$

(P1) If a married woman committed the murder, then if Mrs Peacock didn't do it, it was Mrs White.

(P2) A married woman committed the murder.

(C) If Mrs Peacock didn't do it, it was Mrs White.

However, informational consequence validates all these arguments. To that extent Bledin argues that this conception, unlike diagonal consequence, "coincides with good deductive argument."[43] Bledin also does work explaining away apparent counterexamples to rules that informational consequence validates. K&M argue that Modus Ponens fails within hypothetical reasoning like the following.

1	The streets might not be wet	Premise
2	If it is raining, the streets must be wet	Premise
3	It is raining	Supposition
4	The streets must be wet	2,3
5	It is not the case that the streets must be wet	1
6	$\bot$	4,5
7	It is not raining	3–6

But Bledin rightfully points out that the defender of informational consequence has alternative explanations of what goes wrong in this reasoning that would safeguard Modus Ponens. For example, introducing a supposition in step 3 may consist in provisionally adding information to the 'informational background' relative to which the goodness of an argument can be assessed according to the informational conception of consequence. Against this suppositional backdrop, one cannot safely reiterate the premise 1, which held true relative to a different informational background.[44]

[43] Bledin (2014, 292).

[44] Bledin also considers another way to resist the argument from the perspective of informational consequence, corresponding to a different way of understanding the effects of supposition—see Bledin (2014, 297-8).

Or consider the problems informational modals raise for reasoning by cases (analogous to those we saw deontic modals raise in the miner puzzle).

1	John is in or Niko is in	Premise
2	If John is in, it must be Monday	Premise
3	If Niko is in, it must be Friday	Premise
4	John is in	Supposition
5	It must be Monday	2,4
6	Niko is in	1
7	It must be Friday	3,6
8	It must be Monday or it must be Friday	1, 4–7

Rather than locating failure in the hypothetical uses of Modus Ponens, we can again take on board the idea that a supposition provisionally adds information to relevant informational background. The conclusions under supposition at 5 and 7 are fine. The problem is assuming that conclusions that hold relative to various restricted information states relevant at those lines can be exported to hold of broader unrestricted information states, as the final conclusion at 8 would require.

I am broadly sympathetic with all of Bledin's claims about these examples. In particular, *if* modals and conditionals have an information-state semantics like those we've seen so far, I think Bledin is right that informational consequence seems to correspond much more naturally to intuitively good deductive arguments than does diagonal consequence.

I would of course register a small caveat given the work of §11.2, which is that depending on the framework we use, it is not obvious that what we are tracking good *inference* or deduction here—at least not directly. And I do not think this concern is entirely misdirected. Bledin cites Yalcin (2007) as an inspiration for his view,[45] and we've seen that on Yalcin's view it is important to recognize that modalized contents cannot be used to directly characterize the starting or endpoints of inference. What is more, Bledin repeatedly

[45] Bledin (2014, 280).

characterizes the virtues of informational logic in terms not only of deduction, but (good) deductive *inference* in particular. When defending Modus Ponens he defends the rule as one in which "we can infer" the conclusion from the premises, and even clarifies that he means to "use 'infer' in a thin sense. Inference consists of recognizing what follows; it need not culminate in belief."[46] This is just the sort of thing we would want to say about inference as a mental act insofar as it bridges not only attitudes like belief, but also supposition.

As I say, this is a minor caveat. Even if some frameworks will preclude modalized contents from directly describing attitudes participating in deductive inference, they can do so indirectly as we've seen can be captured by relations of open consequence. There is nothing to prevent someone like Bledin from claiming that our intuitions about good deduction in fact track these broader relations. (Or, at least, this is no less plausible than denying modalized attitudes are new attitude states relating agents to distinctive contents to begin with.) This is not to mention that one could employ a different embedding framework for the information-state semantics on which there are distinctive modalized contents for attitudes.

What is more troubling is that Bledin repeatedly casts the moral of his investigation as undermining a familiar conception of logic according to which it concerns itself with relations of necessary truth-preservation. In the abstract to his paper, he casts the problems for $\models_{Tr}$ in this light.

> Do logically valid arguments necessarily preserve truth? Certain inferences involving informational modal operators and indicative conditionals suggest that truth preservation and good deductive argument come apart.
>
> (Bledin, 2014, 277)

Again, when previewing the importance of the good patterns of deduction which he uses to motivate informational consequence:

> We seem . . . to face a difficulty choice: we can either maintain that these good deductive arguments are valid, and abandon the view that even *material* truth preservation is a necessary condition for validity; or we can maintain that logic is about truth preservation, and undermine the connection between validity and good deductive argument.

[46] Bledin (2014, 284, n16).

(Bledin, 2014, 279)

And again, after reviewing the arguments that raise trouble for $\models_{Tr}$:

> Though the idea that logically valid arguments necessarily preserve truth by virtue of logical form is well entrenched in the philosophical tradition, so too is the idea that good deductive arguments are logically valid. We maintain one of these ideas only by seriously undermining another.
>
> (Bledin, 2014, 290)

In all these passages Bledin is pointing out problems not specifically for the formal relation $\models_{Tr}$, but for an informal, intuitive conception of validity that he characterizes early on in his paper.

> I use the definite description 'the truth preservation view' to denote a cluster of widespread intuitions about the *informal* concept of logical validity. The most basic intuition, of course, is that a logically valid argument with true premisses has a true conclusion. But two further intuitions sharpen this core condition. The first is that validity involves a modal element: it is *impossible* for each of the premisses of a logically valid argument to be true and for the conclusion to be false. The second is that a logically valid argument preserves truth by virtue of the logical form of the sentences in the argument, and not due to the meaning of any non-logical symbols.
>
> (Bledin, 2014, 281, footnote suppressed)

It is clear that Bledin takes the inferences intuitively validated by informational consequence to apply pressure to this view. There is at least one place where Bledin introduces a qualifying hedge that mediates between the problems with $\models_{Tr}$ and the problems for the intuitive conception.

> ...if validity is understood in terms of necessary truth preservation and the formal relation $\models_{Tr}$ explicates this informal target notion, then these arguments reveal that validity and good deductive argument do not line up.
>
> (Bledin, 2014, 289)

The key question, however, is whether $\models_{Tr}$ *does* capture the informal target notion *in the current setting*. Though Bledin signals sensitivity to this issue with the hedge in the above quote, he does not probe the issue in much detail. The closest passages where he speaks to the connection are ones like the following.

> …the relation $\models_I$ does not preserve truth at an index and so is not a possible explication of an informal consequence relation that preserves truth at a context in the ordinary sense.
>
> (Bledin, 2014, 294)

Here Bledin switches to talk about an informal conception of logic as concerned with not truth, but truth at a context. But from the surrounding text it seems reasonable to think the 'widespread' informal conception of logic mentioned before is being rejected.

Why would it matter that $\models_I$ does not preserve truth at an index? Well, one reason it would matter is if the informal notion of logical validity as requiring preservation of truth could somehow be *conceptually tied* to the notion of truth at a point of evaluation. But it is not obvious why we would posit such links. There might be some pressure to think this if we were presupposing a post-semantics that defines sentence truth (at a context, etc.) rendering preservation of truth at an index important for practices of assertion. But not only does Bledin (ostensibly to maintain neutrality) not adopt any such postsemantics, but *even if he had*, it would not yet rule out space to think of the logic as concerned with relations of necessary truth-preservation. For one thing this would still leave open is how the notion of truth at a point of evaluation relates to the truth-conditional structure of *attitudes* like belief and supposition. If connections there are severed, it could likewise sever the connections between the notion of truth at a point of evaluation and the *kind* of truth-preservation that matters to deductive inferential logic as I've been conceiving of it.

In fact, once we think in these terms, we can see that on at least some elaborations of information-state semantics the preservation of truth at a point of evaluation has no direct relevance to the kind of truth-preservation that matters for deductive inferential logic. For example, on Yalcin's view, the world parameter remains completely neutral as we use information-sensitive language to characterize the structure—the *truth-conditional* structure—of a mental state. So if we care about relationships of necessary truth-preservation between

attitude states, then at least sometimes preservation of truth relative to a point of evaluation—which includes sensitivity to the world parameter that is inert in certain attitude ascriptions—is not at issue.

And we can say even more than this. As I stressed in discussing the extension of Yalcin's views in §11.2.1, there is a conception of informational consequence on which it is indirectly tracking relations of necessary truth-preservation through relations like open consequence. Recall that on modest assumptions, informational consequence is coextensive with open consequence, and that the latter is merely one way of grouping classical inference patterns together—inferences which (as argued in Chapter 7) preserve truth necessarily. On this construal, informational consequence is just another way of modeling the kind of necessary truth-preservation that undergirds familiar deductive inferential patterns.

Granted, this extension of Yalcin's framework is idiosyncratic. But a key point of §11.2 is that it is extremely difficult to ask framework-independent questions about the logic for information-state semantics to begin with. In this context, seeing that even one framework allows us to preserve a view on which logic is ultimately concerned with relationships of necessary truth-preservation suffices to show that Bledin's rejection of the latter conception turns on unarticulated and contestable hypotheses about the role of an information-state semantics within a broader theory.

And we needn't stop there. One may be concerned that constitutive relations in the extension of Yalcin's framework are doing all the work of tying informational consequence to the preservation of truth. But we can actually extend these connections to competing frameworks which abandon those constitution relations. To that end, let me turn to a broad class of frameworks (like that sketched in 11.2.2 for MacFarlane) which allow for 'new' attitudes corresponding to modalized and conditionalized language that cannot simply be read off of attitudes taken to ordinary truth-conditional content.

Now how does the conception of inference advocated in Part I get influenced by the inclusion of these new attitudes? Arguably not much. The information contained in an attitude state is still structured truth-conditionally. This is, so far, hard-wired into the semantics for the information-state-sensitive language which is supposed to characterize mental states. Accordingly one can still start by (say) supposing certain premises, and wish to safely extract the truth-conditional information one thereby has supposed. All that has hap-

pened when we include information-state-sensitive language is that we have included language that designates properties of truth-conditionally structured states that do not correspond directly to the acceptance of a truth-conditional proposition—properties that could sometimes be instantiated in multiple incompatible ways. No matter. In any given actual case, a new supposition reflected in the acceptance of a new information-state-sensitive sentence will ultimately have a single effect on the truth-conditional structure of the suppositional state, or it will have no such effects at all. There is just no room in the structure of mentality, so far, for the new supposition to do anything else.

There is, it should be emphasized, one significant complication that information-state-sensitive language has introduced. This is that this language may characterize *more than a fragment of the truth-conditional structure of a starting acceptance state*. Typically, when considering the goodness of an inference involving truth-conditional propositions, we would only need to consider the premise-propositions accepted in their relation to a candidate conclusion. In this way, we might only need to look at 'part' of the information subsumed by an acceptance state (which, after all, may accept many more propositions than just those figuring as premises in an inference). By contrast, coming to accept a new modalized or conditionalized claim may result in a restructuring of a *total* attitude state. As such, to properly track the goodness of an inference described with this language, we are pressured to always consider the truth-conditional information in the *total* information state characterized by the language. Another way of putting things is to say that we must think of the total information in a characterized acceptance state as *all* figuring in the role of a premise.

To see how this shift should be accommodated, we should reflect on what effects arise for the truth-conditional structure of a total state of supposition (belief, etc.) in virtue of its accommodating a new supposition (or belief, etc.) corresponding to the acceptance of an information-state-sensitive sentence. I will make two assumptions about this. First, I will assume that when an agent attitudinally relates to a sentence expressing a proposition, then the information in their acceptance state supports that sentence.[47] Second, I will assume that when one passes from one acceptance state to another by adding a new acceptance, this proceeds roughly along the lines we saw for the 'pro-

[47] In the terminology of §11.2.2: $\{\phi \mid |\phi| \in S_p\} \subseteq S_\alpha$. In other words, I will assume that the aberration that threatened MacFarlane's system doesn't arise.

visional updates' by conditional antecedents. MacFarlane's update below will suit well enough to illustrate the idea.

$$i +_m \phi = \{i' \subseteq i \mid i' \rhd \phi \text{ and } \neg\exists i'' \supset i' : i'' \rhd \phi\}$$

So I will assume that if one starts in a belief state B (say), whose truth-conditional informational structure is given by B_i, then the 'mere' addition of a new belief corresponding to a sentence ϕ results in a new belief state B' such that B'_i is one among the states in $B_i +_m \phi$.

Note that this second assumption presumes a substantive, though natural kind of 'stability' in the truth-conditional effects of old attitudes when 'merely' adopting a new single attitude. To see how this could be violated suppose I have only a single belief: $\Box\phi \vee \Box\neg\phi$ for truth-conditional ϕ and epistemic modal $\Box$. On some views, I might count as believing this by getting into a belief state whose truth-conditional structure contains only ϕ worlds. But imagine I now transition to a 'new' belief state which again involves but a single belief in $\Box\phi \vee \Box\neg\phi$. The only difference is that the truth-conditional structure of the new belief state only contains $\neg\phi$ worlds. While there may be some sense in which I have 'kept the same beliefs' (as I only 'believed one thing' the whole time), the more natural view is that my belief state has shifted in a rather substantial way. I am assuming that in the ordinary case of inference, if one's starting attitude state rules out some worlds, and one maintains all one's attitudes throughout an inferential transition, then one continues to rule out at least those same worlds by the end of the transition. This is why in computing the set $i +_m \phi$ we are justified in considering only subsets of i.

Now, in this context, a good inference would be a transition between *total* acceptance states in which one is guaranteed to rule out no more worlds in the concluding acceptance state than are ruled out by the premise acceptance state because of the relationship between the premises accepted and the resulting conclusion (while this fact is appreciable to the inferrer). That is, it would be a transition in which one is assured of the following: that for any world, if the *sum* of the information in a total starting belief state accepting (and so supporting) the premises is true at that world, then the *sum* of the information in a given acceptance state integrating the conclusion will be true at that world as well.

Note then that even once we add sentences characterizing classes of mental states in the way that informational modals and conditionals do, an interest in

good deductive inference still leads us directly to an interest in relationships of necessary truth-preservation, and for the same reasons they were ever of concern. It's just that now we are tracking truth-preservation for bodies of information that are larger than those given by the premises and conclusion considered in isolation. This is simply because information-state-sensitive sentences require us to take a broader view of the total attitude states in which premises and conclusions figure, which was unnecessary in the case where all sentences expressed only truth-conditional content.

So a good inference, which involves a transition from a state B accepting (and so supporting) premises in Γ to another B' through the 'mere' acceptance of a new conclusion ϕ, will be one that preserves truth relative to the information contained in these states. In other words, it is a transition such that $B_i \subseteq B'_i$. But, of course, slightly more is required for the inference to be an *appreciably* good transition. For it could be that $B_i \subseteq B'_i$ for reasons having nothing to do with the premises accepted, in which case it may be unsafe to *base* the inference only on the premises. To ensure that the premises are doing the work of securing the relevant truth-preserving relations, we want not only that $B_i \subseteq B'_i$, but that *any* attitude state A with information A_i that accepts the premises in Γ is such that $A_i \subseteq A'_i$ for each $A'_i \in A_i +_m \phi$.

Now recall that if $A'_i \in A_i +_m \phi$, by definition $A'_i \subseteq A_i$. So, the condition on preservation of truth $A_i \subseteq A'_i$ is equivalent in this context to the condition that $A_i = A'_i$. But this would mean that $A_i \in A_i +_m \phi$, which holds if and only if $A_i +_m \phi = \{A_i\}$—i.e. if and only if $A_i \triangleright \phi$.

Note what we have just argued. We started by saying that an inference from some belief state accepting (and so supporting) premises in Γ to one 'merely' adding acceptance of ϕ is a good inference just in case it is a form of inference that ensures that the sum of the information in the starting state is true at every world at which the sum of the information in the concluding attitude state is. And we argued that this condition holds just in case an arbitrary state supporting the premises in Γ supports ϕ.[48] And that is, of course, just the condition required for ϕ to be an informational consequence of Γ.

In this way, with a few natural assumptions about how the truth-

[48] Well: an arbitrary *acceptance* state whose truth-conditional content accepts the premises must accept the conclusion. There will be a gap between informational consequence and the conditions on good inference if there are no possible acceptance states corresponding to certain truth-conditional information states (see the similar caveat for the relevance of open consequence in §11.2.1). I assume here that this gap doesn't arise.

conditional information in a total attitude state responds to the integration of new information corresponding to (potentially information-state-sensitive) sentences, we find that a concern with good inference construed precisely as requiring necessary preservation of truth leads directly to consideration of informational consequence. Preservation of support by an information state across premises and conclusion is equivalent to preservation of truth at every world of the information in the state as it shifts from accepting the premises to the accommodation of the information in the conclusion. We quantify over such acceptance states, when investigating good inference, to assure that the preservation of truth is secured by the acceptance of the premises, and not by other features of the state.

In short, on both constitutivist views like that offered by Yalcin, and natural elaborations of non-constitutivist views like that offered by MacFarlane, we find time and time again that the natural formulation of a view of logic as concerned with truth-preservation leads to formalizations using informational consequence. So why do theorists like Bledin think there is such a tight connection between truth-preservation and formalizations like $\models_{Tr}$ rather than $\models_{I}$? There may be a presumption lurking in the background concerning the relationship of the compositional machinery to the logic: a presumption that the world parameter in the compositional semantics has some privileged role to play in assessing relations of truth-preservation. This kind of presumption would underestimate the varying ways that different kinds of world-relativization in the semantic value of sentences can bear on the necessitation relations we care about.

We encountered something like this issue earlier in Chapter 8. There I discussed two competing logics for the same modal language, where both logics were concerned to track truth-preservation across exactly the same range of possible worlds (which could, e.g., have been metaphysical possibilities). But how could there be *competing* logics which agreed on the compositional semantics, the aim of logic as tracking truth-preservation, and even on the range of worlds over which truth was to be preserved? The answer was that the compositional machinery allowed the same possible worlds to play multiple roles in the evaluation of a sentence. Accordingly, it was an open question how to *relate* sentences to the sets of worlds over which it was agreed truth should be preserved. With information-state semantics, we encounter a similar issue on loosely similar technical grounds. Truth-at-a-point-of-evaluation

in information-state semantics are determined relative to both a world and a set of worlds. The worlds that can figure as values of the world-parameter, and those that can figure among the values of the set-of-worlds parameter, can be the selfsame *kind* of words (again, say, metaphysically possible worlds). What this means is that even if we agree that logic should care about necessitation relations across a single kind of worlds, our semantics gives us multiple choices for how to use sentences to investigate those necessitation relations. On the view I favor, which would also be natural given Bledin's stated focus on deductive inference, these relations should be settled based on how sentences characterize the truth-conditional structure of mental states. And this precisely *speaks against* attention to the world parameter of an index in certain cases—namely, those where information-state-sensitive language is at issue.

So I think that Bledin is wrong to think that the viability of information-state semantics creates a tension between an intuitive conception of logic as concerned with truth-preservation and an intuitive conception of logic as tracking good deductive inference. On the contrary, with modest care in handling the application of the semantics, we see that these two intuitive conceptions are as tightly linked as ever.

Having said that, I should concede that Bledin is right to emphasize that *something* important has shifted in the transition from $\models_{Tr}$ to $\models_{I}$ for understanding what a logic for an information-state semantics is or could be. When Bledin rejects the intuitive conception of logic as tracking truth-preservation, he presents what he takes to be the alternative 'informational view' in passages like the following.

> Facts about validity, on this informational view, tell us about the structure of the bodies of information that we generate, encounter, absorb, and exchange as we interact with one another and learn about our world.
>
> (Bledin, 2014, 280)

But this claim could be agreed to by many conceptions of logic, including truth-preservation views. Some truth-preservation views care about truth-preservation precisely because of its role in telling us about the structure of bodies of information like those in mental states. The key question is: *what structure* is validity telling us about? Bledin frames this in terms of deduction.

> Deductive argumentation, on the informal, pre-theoretic picture I have had in mind, is an information-driven enterprise in which

> an agent investigates what is so according to a salient body of information that incorporates the premisses of an argument. In many contexts, this body of information is the informational content of the agent's beliefs—the agent is trying to determine how things are in the actual world. But it need not be. An agent might be investigating what is so according to the clues in the famous zebra puzzle, or a politician's stump speech, or the testimony of an untrustworthy eye witness to the murder, and so on. If this testimony incorporates that Colonel Mustard did it and that if Colonel Mustard did it then he used the candlestick, what else does this information incorporate?
>
> (Bledin, 2014, 303)

Again, it is not easy to discern what in this passage is out of line with a standard view about truth-preservation, like that which I described in Part I. And if what I've just argued is correct, there needn't be any conflict. Still, the passage alludes to important changes in how logic operates. First, there is a focus on *total* information states—'bodies' of information. But it is important to be clear about the contrast: total information states *rather* than the information 'merely' contained in some premises. The reason for this shift is that, with the accommodation of information-state-sensitive language, we cannot speak of the (truth-conditional) information 'merely' contained in some premises. The premises do not on their own determine any unique information of this kind. This is also why Bledin reasonably uses elliptical language describing "what is so" according to information, clues, etc. Ordinarily this would mean: what is *true* according to the information, the clues, etc. But it is not clear we can speak about truth in this context. And even if we could (say, because of a suitable postsemantics) it could be misleading to do so. What we care about, when we care about information-state-sensitive language 'being so' according to a body of information, is for the *total* information state to be structured in some way characterized by that language that need not correspond to any truth-condition.

So there is a substantial shift in logic with the change to information-state semantics. Because information-state-sensitive language can shape the truth-conditional structure of total information states, we cannot concern ourselves merely with the 'truth-conditional content' of premises, as there may be no such content. In assessing relations of truth-preservation, logic must take a

broader view of the relationship between the total truth-conditional information in states merely accepting premises and those which also accept the conclusion. That is why informational consequence is defined with respect to such total states. But the information we are concerned with in this way *is* truth-conditional, and it is *still* the preservation of truth in the transition between states—now total states—that is of concern to logic.

The forced shift to a concern with total states would be an important one for logic. But it is critical to recognize that, in this context, it would represent no shift from a concern with necessary truth-preservation. On the contrary, necessary truth-preservation is as important as it ever was to logic and precisely for our now familiar reason: such preservation is an integral condition on the performance of a good deductive inference.

11.4 Modus Ponens and Weak Belief

Though I've been critical of what I take to be a conflation in Bledin's arguments, I should stress again that I am otherwise greatly sympathetic with his broader methodology and its outcome. I enthusiastically endorse the (cautious) use of intuitions about good inference as a test for the viability of a given formalization of a logical consequence relation. And I accordingly agree with Bledin that if information-state compositional semantics end up being our best semantics for modals or conditionals, *and* the semantics are applied in the ways they popularly are, *then* informational consequence is a better candidate to stand as a logical consequence relation concerned with good deductive inference than is diagonal consequence. My limited disagreement with Bledin, if anything, should actually strengthen that case. Informational consequence can now 'tick all the boxes' and be interpreted to satisfy both the intuitive view that logic is concerned with truth-preservation and the intuitive view that it is concerned with good deductive inference.[49]

Bledin's defense of informational consequence by examining particular candidate inferences is, by his own admission, incomplete and partial. And of course it would have to be. There are too many possible inference patterns

[49] I should flag that I would not want to endorse informational consequence unrestrictedly, though. As alluded to earlier, Mandelkern (2020) makes an important case that informational consequence delivers inappropriate results in the presence of counterfactual conditionals. Sadly I do not have space to discuss the implications of these examples for logical consequence here.

to consider, each of which could be the subject of a lengthy paper. Here I want to circle back to the inference rule with which I began this chapter—Modus Ponens—and in particular to circle back to McGee's apparent counterexamples to it. There are two reasons to reconsider the rule. First, Bledin says some helpful things about McGee's examples, but which I think still fail to constitute an adequate defense of the rule on logical terms. The discussion of how to complete Bledin's case will lead naturally to consideration of probabilities and their connection to logic. Second, this engagement with probabilities will in turn lead us to semantical considerations that actually *could* present quite substantial challenges to the general applicability of the conception of logic which I am putting on offer, and which arise naturally within information-state semantics for modals and conditionals.

Bledin opens his discussion of the problems for truth-preservation views of consequence with a variant on McGee's counterexamples, which we already noted in §11.3.

(P1) If a married woman committed the murder, then if Mrs Peacock didn't do it, it was Mrs White.

(P2) A married woman committed the murder.

(C) If Mrs Peacock didn't do it, it was Mrs White.

Bledin stresses that, in spite of McGee's claims, this is a *good inference*.

> *Pace* McGee, I think that we can appropriately make this argument in both categorical and hypothetical deliberative contexts. Publicly, if someone asserts or supposes that if a married woman did it then if it was not Mrs Peacock it was Mrs White, and also asserts or supposes that a married woman did it, then we can infer on this basis that if it was not Mrs Peacock it was Mrs White. Privately, if you activate your beliefs or simply suppose in an episode of internal theoretical deliberation that the conditional and its antecedent both hold, then you can infer that the consequent holds.
>
> (Bledin, 2014, 283–4, footnotes suppressed)

I agree. But a defense of informational consequence is not really complete until one has explained the relevance of all this to McGee's claims that it is reasonable

for him to believe the premises of his instance of Modus Ponens and disbelieve the conclusion. Bledin seems to do so in a footnote where clarifies what he means when he says that one "can infer" the consequent from the premises. He qualifies:

> This is *not* to say that in categorical deliberative contexts involving assertion and belief activation you should come to believe that if it was not Mrs Peacock it was Mrs White. Perhaps you believe that (P1) holds or that (P2) holds in the face of strong evidence to the contrary. Still, the *Modus Ponens* inference can shed important light on the normativity of your situation—for instance, that you ought either not to believe both [(P1)] and [(P2)], or to believe [(P3)].
>
> (Bledin, 2014, 284,n.17)

While I agree with Bledin's claims about *inference*, I am inclined to disagree with his claims in this footnote and especially its final remark. McGee's example seems to me to show precisely that what is claimed here is false. It is perfectly reasonable for someone to believe McGee's premises and reject his conclusion in the circumstances McGee describes. Such a character is under no rational obligation to change their attitudes. Indeed, I don't think they even have *any reason* to change them. And I think any claim to the contrary cannot be made on purely intuitive grounds as Bledin appears to do—the prima facie case is certainly in McGee's favor. If it wasn't, there probably would have been no force to McGee's cases to begin with.

Note that Bledin here is implicitly appealing in his footnote to the kinds of bridge principles for logic that I rejected in Chapter 3. Once we reject those, we are not compelled to say what Bledin does in this footnote, which is partly reassuring. But we are still left with the burden of saying why someone seems reasonable in holding attitudes to sentences in McGee's cases that logic condemns as contradictory. While our work on the normativity of logic does not obviously force us into saying something incorrect in McGee's cases, it also does not of itself point to a resolution of McGee's examples. For example, we've seen cases like the Preface Paradox, semantical paradoxes like the Liar or the Sorites, and epistemic paradoxes more broadly where it can be perfectly rational for an agent to believe a set of attitudes that is recognizably logically contradictory, *even* to the agent holding the attitudes. If Modus Ponens is valid,

a character in the circumstances McGee describes would naturally end up believing a logical contradiction. But such a character doesn't obviously *seem* to be stuck in a case like the Preface-Paradox or like the the semantic paradoxes, where one is intuitively just 'doing one's best' given limited information in a non-ideal situation. On the contrary, it is not clear there is *any* felt tension in one's accepting the premises of the relevant Modus Ponens inferences while rejecting their conclusions in the kinds of circumstances McGee describes.

But there is something further that is intriguing about McGee's cases hinted at by Bledin's presentation (and discerned by many other philosophers). Note that in the main text where Bledin defends Modus Ponens he stresses that if someone *asserts* or *supposes* the premises one can infer the conclusion. Only in the footnote does he make the claim about *belief*. I suspect this was not incidental. I, following many philosophers, think Bledin's claims about supposition (and perhaps assertion) are intuitively *correct*, in spite of the corresponding claim about belief being incorrect. For example, sticking to McGee's original case, if you suppose *that a Republican will win the election*, and further suppose that *if a Republican wins the election, then if it's not Reagan who wins it will be Anderson*, then you cannot further suppose that *if it's not Reagan who wins, it will* not *be Anderson* without there being a kind of contradictory tension. It is hard to put the feeling here in non-metaphorical terms. But in the case of supposition, it feels like there is 'no room' between supposing the first two premises and the conclusion that *if it's not Reagan who wins, it will be Anderson*.

A point in this vicinity was defended early on in response to McGee by Over (1987). After giving McGee's election example and quoting McGee's claim that we can properly believe the premises but not the conclusion of his Modus Ponens inference, Over says the following.

> The first important point to notice is that McGee speaks of what we *believe* in the above quotation, and not of what we *assume*. In fact, he never speaks of Modus Ponens as the rule which allows us to infer a conclusion ψ from *assumptions* of the form ϕ and $\phi \Rightarrow \psi$, (or from assumptions these forms depend on), and yet this is how the rule is standardly stated in systems of natural deduction. Actually (3) does validly follow given that we *assume* (1) and (2). Having a belief is not at all the same thing as making an assumption, and one respect in which they differ is that a belief

> *may* be suspended or set aside when an assumption by its very nature cannot be.
>
> Suppose you believe that Reagan will win the election, and I ask you to consider what will happen if he does not win. You have no difficulty — you suspend or set to one side (in very informal terms) your belief about Reagan and certain other related beliefs, and then try to see what plausibly follows. But suppose now I ask you what will happen if Reagan does not win the election on the assumption that he will win it. This is puzzling because I am asking you, in effect, what follows from a contradiction. You do not immediately suspend or give up the assumption that Reagan will win the election, in order to decide what will happen if he does not win it, for what in that case would be the point of making the assumption? An assumption is a proposition we hang on to in circumstances like this, and to make the assumption is to agree for a time to hang on to the proposition.
>
> (Over, 1987, 143, footnote suppressed)

I think Over's remarks here hold equally well of supposition—and perhaps supposition corresponds in some measure to what Over meant by an "assumption."

Either way, this points to an intriguing fact. Certain special sets of beliefs that run counter to Modus Ponens do not have a contradictory feel, while the corresponding suppositions do. This tells us that, from an explanatory standpoint, merely rejecting Modus Ponens on the basis of intuitive judgments is about as problematic as merely endorsing it on their basis. What is needed is a commitment about logic *alongside* some plausible explanation of a persistent asymmetry in perceptions of logical tensions between belief and supposition (and perhaps other attitudes).

Intriguingly, we find just this kind of asymmetry for some other 'informational inconsistencies' as well. Consider epistemic contradictions (and their natural connection to Łukasiewicz's principle). As Yalcin (2007) points out, it seems problematic to simultaneously suppose $\neg\phi$ and $\Diamond\phi$, and for it to be embedded in conditional antecedents, and suggests that it is a virtue of informational consequence that it captures this thought:

> …any formal regimentation of the intuitive notion of consequence should substantially track our intuitions concerning

> what follows on the supposition of what. Now suppose that it is not raining. Given that supposition, might it be raining? Obviously not! Hence $\neg\phi$ and $\Diamond\phi$ are incompatible.
>
> (Yalcin, 2007, 1003)[50]

But focus instead on belief. Consider the election circumstances McGee describes. We already noted that it seems reasonable in those circumstances to believe that a Republican (namely Regan) will win. But: is it therefore problematic to believe that a Democrat *could* still win? Does rationality require that one believe that it is (epistemically) *impossible* that a Democrat win? This seems too strong a conclusion to endorse.

Facts not unlike these have inspired Hawthorne et al. (2016) to argue that 'belief is weak' in the following sense: "the evidential standards that are required for belief are very low." In particular Hawthorne et al. argue that "merely thinking that a proposition is likely may entitle you to believe the proposition."[51] Part of the evidence for these claims comes from the fact that sentences like (11) and (12) are not only felicitous, but "[do] not seem to be any kind of admission of irrationality."

(11) I believe it's raining, but I'm not sure it's raining.

(12) I believe it's raining but I know it might not be.

They cite other interesting arguments for this claim, including that "believes" accommodates neg-raising (roughly, "does not believe" conveys "believes not") and that this reveals that "belief" patterns with other 'weak' attitude and speech act verbs.[52]

Hawthorne et al. focus on the weakness of belief to pry apart the evidential warrant required for asserting a proposition and that for believing it. In particular, they are concerned to argue that warrant required for the latter can be substantially weaker than the warrant required for the former. But I am not interested in warrant for assertion here.[53] Nor am I concerned with an even

[50] It is perhaps worth flagging that it is not obvious this argument is given entirely in Yalcin's own voice. He cites it as a *possible* line of objection to diagonal consequence.

[51] Hawthorne et al. (2016, 1394-5).

[52] See Hawthorne et al. (2016, 1399).

[53] Though strong norms for assertion would explain the reasonability of Bledin's earlier claims about what one can safely conclude from *assertions* of the premises in a Modus Ponens inference.

rough specification of the conditions of rational warrant for belief. Instead, I am concerned merely with the 'weakest' version of the claim that belief is weak: "it seems that one can [rationally] believe p even if one has not ruled out the doxastic possibility that p is false."[54] I will assume in what follows that this claim is supported by the examples Hawthorne et al. give. But I should flag that the ensuing discussion is compatible with several ways of fleshing this idea out. It may be that "believes" is ambiguous between a weak sense vindicating the above claim, and a strong sense which does not (as Hawthorne et al. deny). It may be that "believes" is context-sensitive, with at least one weak interpretation. It may be that "believes" semantically expresses only a strong sense, but pragmatically implicates a weaker attitudinal relation, where our judgments of the rationality of belief sometimes track warrant for the implicated state. It should be possible to reformulate most of what I say below to fit any of these approaches.[55]

The next important claim is that supposition is *not* like belief in this regard: one cannot [rationally or otherwise] suppose p unless one has suppositionally 'ruled out' the possibility that p is false. This is supported precisely by the fact that Łukasiewicz's principle seems to hold without exception in suppositional environments. As soon as one supposes p, against the backdrop of this supposition it is incoherent to suppose that p remains possible. Alternatively: against a supposition that p, p *must* be true.

If something like this story lies behind the asymmetries between belief and supposition, a natural path opens to understand how to understand the implications for logic of the diverging judgments about the goodness of Modus Ponens, and Łukasiewicz's principle.

If it can be true that an agent believes that ϕ (for truth-conditional ϕ) without their doxastically ruling out some worlds where ϕ is false, then *a gap opens up between the content of belief-report complements and the structure of the belief states they characterize*. In particular there is a gap between, on the one hand, the truth-conditional structure of the content of ϕ as used in a true ascription of belief to an agent and, on the other, the underlying truth-conditional structure of a doxastic state in virtue of which the agent is truly said to believe ϕ. This gap opens up space for the following possibility: that an agent

[54] Hawthorne et al. (2016, 1396).

[55] One view which, as far as I can tell, *may not* vindicate the ensuing discussion is the context-probabilist proposal of Moss (2019). I'll come to a discussion of context-probabilism and its relevance to logic soon.

gets into a doxastic state in virtue of which they are said to believe a series of truth-conditionally contradictory contents (contents which collectively rule out all worlds), even though the underlying doxastic state has consistent truth-conditional structure (that is, structure in which not all worlds are ruled out).

Obviously, this will have implications for the rationality of getting into such states. *Even if* there were indefeasible rational pressure against getting into a doxastic state with inconsistent truth-conditional content (which in Chapter 3, essentially following Harman, I cautioned against), that ban would not have any applicability to the aforementioned special class of doxastic states in virtue of which contradictory contents are truly reported to be believed. There should be no logical rational pressure from considerations of coherence against holding the relevant beliefs together, and they should probably not even 'feel' as though they are contradictory in character.

It is reasonable to think this is witnessed quite directly for Łukasiewicz's principle. One can rationally believe ϕ is false without believing it is (epistemically) impossible for ϕ to be true. And the reason this seems unproblematic is precisely that in believing ϕ false, given the weakness of belief, one need not yet have epistemically ruled out that ϕ is true. When I believe in this way, my beliefs do not feel contradictory because, fundamentally, there is no contradiction in my underlying doxastic state, even if there is contradiction between the contents I count as believing in virtue of being in that state.

It is also possible that this situation arises in the Preface Paradox.[56] When one believes each of a series of contents $p_1, \ldots, p_n$, one might count as doing so even though one's underlying doxastic state leaves open 'remote' possibilities at which each is false. But individual remote possibilities might not *collectively* be remote. Accordingly, the same underlying, logically coherent doxastic state in virtue of which one believes each of the $p_1, \ldots, p_n$ might equally be a state in virtue of which one believes the general claim that one among these propositions is false. It is worth adding that even if one can rationally believe what is expressed by each sentence in a long book, while also believing the general claim that at least one sentence in the book must be false, if one *supposes* all these things one appears to have supposed a contradiction. So, revealingly, we seem to find the same doxastic/suppositional asymmetries in the Preface Paradox that we do for information-theoretic language.

[56] I should note that this is not *required* for the defense of informational consequence I am mounting.

Finally, and most importantly, the phenomenon is also plausibly underlying examples like the election case given by McGee, thought this can be tricky to see. Remember that in the election scenario described, one believes a Republican will win because one believes Reagan will win. But *both* of these are weak beliefs. One certainly allows that it possible both that Reagan loses and also that a Republican loses. And the belief has to be weak, otherwise the second premise of McGee's argument and its conclusion would be either defective or only vacuously true relative to one's belief state. One would believe Anderson wins if Reagan doesn't (and that Anderson wins if neither Regan nor a Republican does) only in the sense that one believed Regan wins if he doesn't. The conditional antecedent would throw away all worlds compatible with one's belief state.

Once we recognize the weakness of both the belief that Regan will win and that a Republican will win, we see that the premise *if a Republican doesn't win,…* restricts our attention to a *proper* subset of the worlds compatible with the relevant belief state. It says of those worlds that the non-Reagan-winning worlds are Anderson-winning worlds (which is correct). But the conclusion says that all non-Regan-winning worlds in the doxastic state (which because of the weakness of belief include Democrat-winning worlds) are Anderson-winning worlds, which is false. So it is precisely the weakness of the simple premise belief that a Republican will win that allows the conditional premise to be rationally accepted while the conditional conclusion is rationally rejected. And it does this precisely by facilitating the truth-conditional coherence of the overall belief-state in the face of its being related to otherwise contradictory contents.

It might be easier to see the underlying problem here with a much simpler 'counterexample' to Modus Ponens. Suppose a family member buys me a ticket to a lottery with an absurdly large number of tickets. It is perfectly reasonable for me to believe that my ticket will lose—that is, that a ticket other than mine will be the winner. But I should also believe (indeed, be as confident as I am in the existence of the fair lottery) that *if* a ticket other than mine is the winner, it is not possible that my ticket wins. On the assumption that my ticket loses, it is incoherent to suppose it could win. But it is not rational for me to conclude from all of this it is not *possible* that my ticket wins. It is manifestly possible—it is just that it is a highly remote possibility.

What we see in this case is a contrast between the weakness of my belief

that my ticket will lose, and the corresponding 'strength' of the conditional update. The conditional antecedent rules out worlds that are not ruled out by my doxastic state, even though the state and the antecedent are characterized using exactly the same sentence. A conditional antecedent seems (as the informational semantics would have it) as strong as a supposition. This is unsurprising, given the strong connections between conditional antecedents and acts of supposing.

This account of McGee's cases raises a question: if conditional antecedents are always strong, and belief is sometimes weak, why do all the apparent counterexamples to Modus Ponens always involve modalized or conditional language in the consequent of the conditional premise? Shouldn't we find similar apparent counterexamples where conditionals don't embed other conditionals or modals?

The answer is that we do not find these because though conditional antecedents 'contravene' the weakness of belief, conditional *consequents* need not. I can rationally believe that *if* my relatives buy me a single lottery ticket, that ticket will lose, without being rationally compelled to believe that if my relatives buy me a single lottery ticket, it is not possible for the ticket to win. Accordingly when one infers, under weak belief, ψ from *if* ϕ*, then* ψ and ϕ, for ordinary truth-conditional ϕ and ψ, the inference is bound to feel good (in fact it *can* be good) even though one's belief in ϕ leaves open some 'remote' non-ϕ worlds. Granted, believing the conditional *if* ϕ*, then* ψ only requires a proper subset of one's doxastic alternatives—the ϕ worlds—to be ψ worlds. But if one is rational, that will be enough to rationally compel one to the belief that ψ—that is, the *weak* belief that ψ. One's rational state may, compatibly with believing the premises, leave open some not-ϕ and not-ψ worlds. But for the same reason that the presence of these worlds needn't interfere with believing that ϕ, they also need not interfere with believing that ψ.

Accordingly we should only expect to find apparent counterexamples to Modus Ponens when there is language that 'strengthens' in a conditional consequent by forcing the consequent to characterize the total truth-conditional structure of a belief state. Conditionals do this (through their antecedents, which rise to the strength of suppositions). And epistemic modals have that effect as well: believing p allows for some doxastic not-p alternatives. But believing *must* p does not. And if *might* is the dual of *must*, it will similarly force a characterization of total attitudinal structure.

If this explanation is on the right track, what would we learn about a logic for deductive inference? Well, what we would be seeing is that there can be a slight gap between the truth-conditional structure 'strictly' foisted on an information state by a sentence that one is correctly reported to believe, and the actual truth-conditional structure of the doxastic state that is required for the belief report to be correct. In this setting, as I say, there can be 'contradictory' sentences—sentences which collectively rule out every possibility—which can be rationally believed, because in rationally believing them one doxastically does not rule out every possibility.

Were this the correct explanation of matters, then I think it should be clear that the most perspicuous logic for correct deduction would remain something like informational consequence, and Modus Ponens should be retained as a valid rule of inference—at least if generalized conjunction introduction is retained as well. It is no count against a logic which pronounces certain sets of sentences to be contradictory, that one could consistently and rationally bear certain attitudes toward them precisely because in doing so the truth-conditional structure of those attitudes need not match the structure given by the sentences themselves.

It is also worth noting from this perspective that it is extremely dangerous for the theorist to try to assess correct deductive relations by looking at rational *belief*. Intuitions about the goodness of deductive inference under supposition represents the far clearer case, since it dispenses with the need to carefully track the slack between the truth-conditions expressed by the complement clause in a belief report and the truth-conditional structure of the doxastic state thereby characterized.[57] This provides one final bit of sharpening for the criticism of treatments of the normativity of logic in Chapter 3 that focus on belief to the apparent exclusion of supposition. There, I claimed it was problematic that virtually no logico-normative bridge principle was extensible in any straightforward way to the case of supposition, as supposition states have a standing equal to belief states as subjects of logical normativity. Now, if anything, supposition states appear to be *privileged* with respect to belief states in this respect.

Sometimes reflections on informational consequence like the foregoing

[57] It is in some sense equally important for agents to be cautious about when they infer from their beliefs. But there is no reason to think thinkers have any trouble doing this. I don't think anyone would be accidentally be taken in by a case like McGee's counterexample, or my simpler one, and infer in ways that were impermissible due to the presence of weak beliefs.

have led theorists to say that informational consequence is, or appears to be, the logic governing 'full acceptance.'[58] Saying this is fine as long as what is understood is that, from the perspective of developing a logic of deductive inference specifically, it is not clear what yet could count as relevant to good deduction beyond full acceptance. Even in the examples above where there are inferences involving weak beliefs (e.g. what appears to be a Modus Ponens inference involving such beliefs), the goodness of the inference hinges only on necessary preservation of truth among the worlds compatible with an underlying acceptance state. It's just that sometimes these relations are not easily read off of the sentences that one truly accepts as one undergoes the inferential transition. In particular, it is not clear (yet) that there is any room for a 'logic of partial acceptance,' where the logic tracks the goodness of an inferential relation. Work would have to be done to make sense of that notion.

This as it happens is precisely our next, and final, question: *can* this be done?

11.5 Deduction in the Context of Probabilistic Mentality

In §11.2.1, I flagged that Yalcin's structuring of mental states with sets of worlds was provisional, and awaited refinement to cope with probability modals like "probably" and "likely". He sketches his approach to these modals as follows:

> The basic idea of the approach I want to recommend is simple: just upgrade the kind of object the information parameter can take as a value, from a set of worlds to a *probability space*. The intension of a sentence, relative to context, will be a function from world-probability space pairs to truth values. We will take it that a probability space P determines a probability measure Pr_P over sets of possible worlds …
>
> (Yalcin, 2007, 1015)

More specifically, we take a probability space P to be a triple $\langle \Pi_P, \pi_P, Pr_P \rangle$. Π_P is a partition of the space of possible worlds (I will assume metaphysically possible worlds), which intuitively determines the subject matter or questions to which an information state is responsive, and so determines which propositions the probability measure of the space is defined over. We say that Π_P

[58] See Mandelkern (2020), Santorio (2022).

classifies a truth-conditional proposition p just in case for every cell $\iota \in \Pi_P$: all worlds in ι are p-worlds or all worlds in ι are $\neg p$ worlds. π_P is a subset of Π_P, whose cells give the 'live' possibilities according to the measure.[59] Pr_P is a probability function such that

(i) $\forall \iota \in \pi_P, 0 \leq Pr_P(\iota) \leq 1$, and

(ii) for all truth-conditional propositions p:

$$Pr_P(p) = \begin{cases} \sum_{\iota \subseteq p} Pr_P(\iota) & \text{if } \Pi_P \text{ classifies } p \\ \text{undefined} & \text{otherwise} \end{cases}$$

As Yalcin suggests, in our semantics we have probability spaces occupy the place of the information parameter. It will also be expedient to lift our definition of truth at a world/information-state pair to a definition of truth at a cell/information-state pair as follows.

$$[\![\phi]\!]^{\iota,P} = t[/f] \Leftrightarrow \forall w \in \iota : [\![\phi]\!]^{w,P} = t[/f]$$

The measure Pr_P is exploited in the semantics of $\triangle$ [i.e. 'probably'] as follows:

$$[\![\triangle\phi]\!]^{w,P} = t \Leftrightarrow Pr_P(\{w \mid [\![\phi]\!]^{w,P} = t\}) > \tfrac{1}{2}$$

Just as we saw that epistemic possibility modals delivered a property of information-states construed as sets of worlds (e.g., that they contain some prejacent-worlds), probability modals deliver a property of information-states construed as probability spaces. "Likely" delivers the property that is true of a probability space just in case it assigns probability of greater than $\frac{1}{2}$ to the modal prejacent.

Extending the framework in a straightforward way to attitude reports, Yalcin recognizes, requires construing attitude states as subsuming this kind of probabilistic structure:

> [I]t is natural to conjecture that the semantics for acceptance attitude verbs ('believes', 'knows', 'accepts', 'supposes', etc.) can straightforwardly mirror our earlier domain semantics …Let these verbs shift the value of the information parameter to the

[59] Note: to be 'live' is not necessarily to have non-zero probability, so as to allow for epistemically possible zero probability events.

> information state corresponding to the attitude state of the subject, and let the whole ascription require, for truth, that the complement of the verb be accepted with respect to that information state. The information parameter ranges over probability spaces, so the semantics assumes that these attitude states can be modelled by such spaces.
>
> (Yalcin, 2007, 1017–8)

Yalcin elaborates and refines this view in a number of papers,[60] and the core ideas have been enthusiastically taken up, with adjustments, by a number of other theorists.[61]

The assumption that Yalcin appeals to about attitude states—that the information contained in them can be modeled with something resembling probabilistic or other graded structure—is an extremely popular one. If anything, the idea that 'full' attitudes like belief ultimately give way to degrees of belief or credences is the dominant view in contemporary philosophy of mind. The view is taken to be a natural, if not inevitable, refinement of more familiar truth-conditional or propositional structuring of the information in mental states.

From the logical perspective, it may initially seem that little changes in the shift to probabilistic structuring of mentality. After all, we can easily and naturally extend information-state conceptions of consequence, like informational consequence, to the probabilistically structured case. We do this by first extending the notion of *support*, for example as follows.

> A probability space P *supports* ϕ, written $P \triangleright \phi$, just in case
>
> $$\forall \iota \in \pi_P : [\![\phi]\!]^{\iota,P} = t$$

We can then retain a conception of consequence as tracking preservation of support.

How should we interpret these kinds of changes within the framework for an inference-based logic? The answer is that it is extremely hard to see, because of a surprising gap in the literature on graded attitudes that is just now beginning to be filled. Namely, at present, we have no clear model of what it is to *reason* with such attitudes, let alone to infer with them.

[60] Yalcin (2011, 2010, 2012a).

[61] See e.g. Rothschild (2012), Moss (2018).

An important lesson of §§11.2–11.4 was that as long as one doesn't change the underlying truth-conditional structure of mentality, relatively little can change for our understanding of good inference, and so little of *fundamental* importance changes for logic. Granted, tracking good inference can be made more complex by the introduction of language that characterizes the total informational structure of an acceptance state (say). But these difficulties may have little bearing on how we view actual good inference. For example, even though we saw the possibility of using informational consequence to model a form of open consequence in §11.2.1, we also saw that in application to Yalcin's framework this ended up being a *re-description* of modal- and conditional-free inference patterns—perhaps simply classical ones.

The importance of probabilistic mentality is that it could teach precisely the complementary lesson: that as soon as we restructure mentality, this can have substantial implications for our understanding of inference, and so too for an inference-based logic. My goal in the remainder of the section is to briefly get this issue on the table and to sketch some salient ways that an investigation into reasoning with credences could pan out for an inferential logic.

As I say, the key issue is that it is unclear what it would be to *reason* with probabilistically structured mental states. The natural place to turn to understanding reasoning with graded attitudes would be the subjective Bayesian tradition in formal epistemology, which contains an extensive exploration of the norms governing partial beliefs or credences. The problem is that Bayesian frameworks tend to be applied in highly idealized settings, where theorists have abstracted away from the kind of structure that would be relevant to ordinary human reasoning.

The subjective Bayesian posits two norms governing credal states. The first norm, Probabilism, is a synchronic norm that a rational agent's credences should conform to standard Kolmogorov probability axioms. The second norm, Conditionalization, is a diachronic norm that says that a rational agent will respond to the acquisition of new evidence by conditionalizing on it.

It is worth stressing that when these norms are presented, they are often qualified as norms of *ideal* rationality, or norms governing ideally rational agents. And this is important, because it is not obvious that either norm could provide standards of reasoning for any ordinary human reasoner. In fact, it is not obvious that either norm could provide standards for *reasoning*, in any familiar and recognizable sense, at all.

For example, the first constraint given by subjective Bayesianism is a synchronic one. As I stressed from various angles in §3.4 and §5.1, inference and reasoning more broadly are activities that take time. Probabilism requires an agent to exhibit credal omniscience with respect to all (at least classical) logical necessities. But it doesn't give us a rational means to *arrive* at the credences in question.[62]

Conditionalization has the right relational form to govern reasoning. Here the main concern is that it cannot *exhaust* our account of reasoning with credences. First, Conditionalization is a rule about how credences should adjust in response to *new evidence*. But not all forms of reasoning, including those we would expect to find with credences, should have that character.[63] Indeed, much if not all logical reasoning should intuitively be possible against a fixed backdrop of evidence or no evidence at all. Second, conditionalization defines an update which takes one probability function to another, and so appears to be an update between *total* credal states. There are worries about whether this could ever provide a realistic model for reasoning employed by humans, owing to concerns of computational complexity.[64] And even if it could model some human reasoning, it also seems like we should allow for *some* forms of reasoning with credences that would not involve operations over a total graded state.[65]

These concerns are of course merely *prima facie* challenges. Still, the point remains that non-trivial work would need to be done to defend a Bayesian framework as a framework for reasoning (let alone inference), or to elaborate its transformation into one. So far, detailed work of this kind has yet to be accomplished.

[62] Compare related complaints in Dogramaci (2018) that Bayesianism cannot supply us with a plausible account of (the credal analog of) doxastic justification, even if it could perhaps provide us with an account of (the credal analog of) propositional justification. Though I frame the issue in terms of the passage of time, it is perhaps worth noting that this may not be the most fundamental concern. There are important arguments that rational norms only apply to an agents at a single time—see Hedden (2015b,a). While reasoning presents a challenge to this thesis, it is not obviously a dispositive one (see Hlobil (2015), Podgorski (2016) for some discussion). But even if norms only apply to an agent at a time, we need some understanding of how attitudes are rationally *based*. So we might at least say: reasoning is *relational* and the synchronic norm given by Probabilism lacks this important relational structure.

[63] See , 4542 (crediting Alan Hájek for the point), Dogramaci (2018).

[64] Harman (1986, 25–6), though see Staffel (2013, §3.2) for a reply.

[65] In connection with this point, Hlobil (2016a) notes that it is surprisingly hard to make sense of how there could be *chains* of reasoning in a broadly Bayesian framework, in which an intermediary conclusion is taken as a premise in a new bit of reasoning.

Rather than trying to extract a framework for reasoning directly from the Bayesian view, it is perhaps more profitable to start fresh and try to build a framework for reasoning with credences from the ground up. And there are some natural routes to explore along these lines. Let me outline two.

A first view that easily accommodates reasoning with credences or partial beliefs is a view on which credences *reduce* to 'full' acceptance states with ordinary truth-conditional content concerning probabilistic information of some kind.[66] On this view, fundamentally there are only 'on/off' acceptance states like belief that take truth-conditional contents as objects, though sometimes the condition for the truth of the relevant content is that some probabilistic structure or relation be instantiated in the world. To take a simple example, perhaps having high credence in the proposition *that it will rain tomorrow* is just to have a full belief in the distinct proposition *that it is likely that it will rain tomorrow*.

But what would the truth-conditions of a claim *that it is likely that it will rain tomorrow* be? A natural idea, though perhaps not the only one, is that these could involve *evidential probabilities*—probabilities that are relativized to a given body of evidence.[67] In this case, the evidence that is relevant might be that possessed by the agent to whom we are attributing probabilistic beliefs, or evidence easily available to them, or perhaps evidence they suppose into existence when they suppose how things are, and so on. So on this view, having high credence that it will rain might just be to believe that there is evidence about for rain (storm clouds on the horizon, for example). Whether there is evidence of this kind is an ordinary truth-conditional matter.

However we flesh out this idea, we will get the possibility of reasoning and even inferring with credences. This is simply because reasoning with credences is nothing other than reasoning with the ordinary attitudes of acceptance to which they reduce. There may be distinctive aspects of this reasoning connected with necessary, appreciable truths about probability. For example, it might be a necessary, appreciable truth about evidential probabilities that they obey the Kolmogorov axioms. As such, one could deductively infer from a

[66] See Lance (1995), Schiffer (2003, 200), Holton (2014) for some reductivist sympathizers. For some further discussion see Weisberg (2013) and Dogramaci (2018), and for some linguistic motivations for the view, see the descriptivist alternative explored in Marushak & Shaw (ms./2020).

[67] See Plantinga (1993, chs.8,9) (who uses the term "evidential probability") and Williamson (2000, ch.9).

credence of .7 in a sunny day to a credence of .3 in a sunless one simply by deductively inferring from the claim *that there is a 70% chance of sun* to the claim *that there is a 30% chance of no sun*.

I mention this conception of credal inference mostly to set it aside. The most perspicuous 'logic' in this setting would be relatively boring. We would start with a language initially free of probabilistic talk, and then augment it with probabilistic vocabulary usable to state ordinary truth-conditions that also happens to be taken as 'logical.' The result would be a truth-conditional logic for the word "probable" (say), not substantially different in principle from other forms of epistemic logic. Not only is does the position have little distinctive to teach us about logic, but the view accords poorly with the kinds of motivations that have been adduced for information-state-semantics for probability modals like those offered by Yalcin.[68,69]

So let's instead suppose that reductivism does not hold, and that there are graded attitudes like credence that do not reduce to truth-conditional ones. How might we understand inference to apply to such states?

[68] For example, the view on offer threatens to run headlong into problems with 'higher-orderism' about attitudes, and it fits poorly with expressivism about probabilistic information—see Yalcin (2007, 2011) for early discussions of both ideas.

[69] Dogramaci (2018) develops a framework for reasoning with credences which, while not assuming reductivism, assumes a very closely related thesis we might call 'parallelism': that credences and full beliefs about evidential probability necessarily occur together. In particular, Dogramaci endorses the following principle:

(CC) Necessarily, an attribution of a credence is correct if and only if a corresponding attribution of a belief about [evidential probability] is correct.

I confess that I am confused, once this kind of thesis is adopted and the associated beliefs are treated as genuinely truth-conditional beliefs, why we would need *distinctive* rules for reasoning with credences (though I do sympathize with Dogramaci's claim that the Bayesian does not supply us with them). Given (CC), if there are rules of reasoning, it would seem there would be rules enough among those for full belief (supplemented with conceptual truths about evidential probabilities). Additional rules for credence would seem superfluous. Nevertheless, Dogramaci develops a series of rules for credence of which the following is representative.

(R2-a) If you know p and q are exclusive, and you have a rational credence of x in p and a rational credence of y in q, then you are defeasibly permitted, on that basis, to adopt a credence in their disjunction equal to the sum $x + y$.

An important reminder here is that I am not concerned with reasoning broadly construed, but with inference. Accordingly, *even if* these were correct rules of reasoning, they wouldn't address the question we need to answer in this section. (R2-a) is weak and defeasible in ways that I argued in Chapter 3 reveal its unsuitability to characterize something like a norm of logic. (This, of course, is no objection to the rules as satisfying Dogramaci's aims—there are simply two different projects here.)

In Chapter 2, I highlighted two key attributes of good inference: reliable preservation of information and appreciability. I will presume for now that a logic governing graded inference would continue to abstract from appreciability, and so will provisionally set that second condition aside. As to information-preservation, the nature of this property was determined by the structure of the information in the mental states inference mediates between. I noted in Chapter 2 that these acceptance states are traditionally modeled with truth-conditional structure. Accordingly, I took information-preservation to be preservation of truth-conditional information in particular. The natural question to ask here is whether truth-conditional information can be 'swapped out' for probabilistic information, so that there could be a mental operation devoted to the preservation of probabilistic information for graded states.

It appears there is no obstacle to characterizing this kind of operation at an abstract level. If probabilistic structure (of any kind) is given by something like a probability space, then information about probability would be given by a set of such spaces—the spaces compatible with the information. A transition between two pieces of probabilistic information is probabilistically information-preserving just in case every probability space compatible with the first piece of information is compatible with the second. Alternatively: every probability space that is a member of the first set of spaces is also a member of the second. Just as necessary preservation of truth is a necessary condition on good deductive inference in the context of full acceptance states, probabilistic information preservation would be a necessary condition on good deductive inference in the context of probabilistically graded states.

One more detail is needed to fill out this picture. In the case of full acceptance states, the question arose: necessary preservation of truth for *which* modality? In Chapters 4–5, I defended the claim that the modality relevant to good deduction is metaphysical. Here we face a similar question: what kinds of probability spaces are relevant to probabilistic information-preservation for good deduction? Let me assume for the moment that the kind of probability that we are considering information *about* is itself probabilistically coherent, in the sense of obeying the standard probability axioms over a state space of metaphysical possibilities—we'll soon turn back to consider what could justify this assumption. I will also assume that in a graded inferential transition the partitions Π_P and π_P don't change (such changes would more naturally fall into the class of 'reasoning broadly construed'). Inferential transitions would

effectively have to be relativized to such a partition.

So from here on I'll assume that probabilistic information is given by (or at least determines) a set of probability states over a single partition type $\langle \Pi_P, \pi_p \rangle$ of metaphysically possible worlds, and that subsume probability functions that are probabilistically coherent. I will call such states *CM-probability states* for short ("C" to mark coherence, "M" to mark the space of metaphysical possibility). And we can say that a transition between one piece of probabilistic information S (given by a set of CM-probability measures) to a second piece of probabilistic information S' is *CM-probabilistically information preserving* if every CM-probability state in S is also in S'.

CM-probabilistic information preservation could give us a necessary condition on good inference for probabilistic mental states just as metaphysically necessary truth-preservation does for good inference. But making sense of such a condition is not of much use unless the structure of mentality is suitably refined to allow for substantive transitions that instantiate the relation in question.

For example, let's suppose that a probabilistically graded state, like a credal state, is modeled by a single CM-probability space, which is essentially the structure given in Yalcin's discussion of mental structure above. In this case, *no non-trivial transition between such states would be CM-probabilistically information-preserving*, as every such transition would be between one CM-probability state and another non-identical one. Could we get around the problem by allowing states to be modeled with spaces containing total probability functions that are *not* coherent (including by not being over metaphysical possibilities)? This would allow for a single type of CM-probabilistically information preserving mental state transition: that from any given mental state modeled using a non-coherent probability space to any other. But this hardly seems like an improvement.

It is worth nothing that even on this picture, Yalcin could be entirely right about the semantics of probabilistic language. And we could develop a formal relation, like information consequence, to relate sentences of the language in the ways we've done before. The point is that without adequate restructuring of mentality, such a logic could not yet model an inferential transition, as we have not given mentality enough structure to witness the transitions.

What kind of structure or refinement would help? There is a loose analogy between our current troubles and the problem of 'logical omniscience' when

modeling an acceptance state with a set of metaphysically possible worlds. Familiarly, this latter picture leads to the view that an agent accepts all metaphysical necessities. Taking a mental state to be modeled by a CM-probability space similarly seems to be modeling the agent bearing the state as 'logico-probabilistically omniscient.'[70]

To begin to introduce the refinement needed, let's assume that an agent's mental state is *relational* in the sense we saw discussed by MacFarlane in §11.2.2. That is, let's assume that there is a set of probabilistic contents or propositions, each of which is, or determines, a set of CM-probability states. We can further assume that we have a language (perhaps formal and stipulated) with sentences ϕ that express these propositions. As before, we will denote the proposition expressed by a sentence ϕ as $|\phi|$. Finally, we will assume that a graded mental state is modeled by a set of such probabilistic contents.

Now would we have enough structure to generate credal inference? Not quite. At this juncture we recapitulate the lessons of §11.2. For we must ask: does the relational structure of a graded mental state *derive* from more fundamental, simpler probabilistic structure?

We could easily have the following view. A mental state is fundamentally characterized by a single CM-probability state P. Then whether an agent's graded mental state relates them to a given proposition $|\phi|$ is determined entirely by whether P is among the states compatible with the information in $|\phi|$. For example, consider the sentence $\triangle\phi$ for ϕ expressing a truth-conditional proposition p. This sentence would express the proposition $|\triangle\phi|$ given by the sets of CM-probability states with functions that assign p a value greater than or equal to .5. It may be that my underlying credal state is given by a single CM-probability state P such that $Pr_P(p) = .62$. Accordingly, it would be true to say of me that my credal state accepts $|\triangle\phi|$, but merely because $Pr(p) = .62$ and $.62 \geq .5$, and not because of any *added* structure reflected by the proposition $|\triangle\phi|$ itself.

It should be clear that on this view, nothing has changed fundamentally about the character of mentality. Accordingly, we still have no room for the possibility of something like credal inference. All changes in acceptances of probabilistic propositions are underwritten by changes in an underlying

[70] Note that important aspects of this problem persist if we instead model a mental state with a *set* of CM-probability spaces, to represent 'mushiness' or indifference, as each member of the set is also probabilistically coherent.

state characterizable by a single CM-probability state. And there are again no non-trivial transitions between single states of this kind that could be CM-probabilistically information-preserving.

What we have in this case is a generalization of the issues that arose for Yalcin in §11.2.1. Recall that although we could give some sense to the utility of informational consequence for his framework, it required treating some sentences related by informational consequence *not* to characterize a condition on good deduction. Instead informational consequence sometimes merely exhibited relations of constitution. Still, even in that case we saw some residual point to using informational consequence in the study of inference. After all, sometimes informational consequence could capture *classes* of good inference with the help of information-state sensitive language, even though the classes seemed not to reveal any informative underlying uniform bases for good inference. But in the probabilistic setting we are examining now we cannot assign informational consequence even a limited role of this kind. In the truth-conditional setting, information-state sensitive language could sometimes characterize a mental state in virtue of its relations to truth-conditional content that could figure in an inferential transition. But in this case, the probabilistic language *only* characterizes probabilistic states, none of which can yet figure in productive inference. So unless we make further changes to mentality, the most perspicuous inferential logic for this framework seems to be one that simply excludes the probabilistic vocabulary. There is not even a limited role for probabilistic language to 're-describe' classes of inference via informational consequence.

We can avoid this problem by following MacFarlane in the move that distinguished him from Yalcin, even were the latter to adopt a relational framework for mentality. This is to take the relational properties to provide *fundamental*, underlying mental structure. By doing this, we allow attitudinal relations to probabilistic contents to 'float freely' of each other, in ways that our foregoing approaches did not allow. This opens up the possibility for various kinds of 'probabilistic irrationality' (analogous to the kinds of akrasia MacFarlane noted arise for him, but not Yalcin). And, related to this, it opens up the possibility for genuine inferential credal transitions.

So now let us suppose that a mental state is given, *fundamentally*, by a set of probabilistic contents (where each content determines a set of measures). At this juncture, we have the possibility of all sorts of transitions between

graded mental states that can be CM-probabilistically information-preserving. To make some of these explicit, it will be helpful to augment our language with a set of sentential operators of the following form (for different values of $0 \leq \mathbb{n} \leq 1$).

$$[\![\mathbb{n}\phi]\!]^{w,P} = t \Leftrightarrow Pr_P(\{w \mid [\![\phi]\!]^{w,P} = t\}) = n$$

Then the following sentences are related by informational consequence, and could in principle correspond to good CM-probabilistically information-preserving transitions between the contents that the sentences express.

$$.\mathbb{7}\mathbb{5}\phi \models_I \triangle\phi$$

$$\mathbb{1}\neg(\phi \wedge \psi), .\mathbb{2}\phi, .\mathbb{3}\psi \models_I .\mathbb{5}(\phi \vee \psi)$$

$$.\mathbb{4}\phi \models_I .\mathbb{6}\neg\phi$$

That is, if one has sufficiently high credence in ϕ by relating to the probabilistic content $|.\mathbb{7}\mathbb{5}\phi|$, one can transition to a *new* state including the content that ϕ is likely. Critically, being in a mental state with $|.\mathbb{7}\mathbb{5}\phi|$ as its content does not of itself ensure that one's global state also takes $|\triangle\phi|$ as a content as well. It is fully possible to 'add' this latter content to one's graded mental state. And the resulting transition would be CM-probabilistically information-preserving. Similarly we can get genuine CM-probabilistically information-preserving mental state transitions corresponding to the second and third relation above as well.

In this way, we finally secure the minimal mental structure seemingly required for probabilistically graded inference, and so for the very possibility of a logic that can help track good inferences of that kind. This is progress. But it is sadly only the first step in filling out an inferential logic for graded states. Let me sketch four issues that remain to be addressed before turning back to consider some general lessons from consideration of probabilistically graded mentality.

First, of course, I have set aside the issue of appreciability. Every transition from no probabilistic premises to certainty (i.e. credence 1) in a metaphysical necessity will satisfy the condition of CM-probabilistic information-preservation. Obviously, even if there were graded inference, not every one of these transitions would count as a good inference for an ordinary human reasoner. The natural refinement is to wheel back in a condition of appreciability to help explain what else would be needed for a good inference to be

performed. I see no reason to think this relation would be any less psychologically variable than in the case of full attitudes. So we should not expect a theory of appreciability to fix some stable set of rules that dictate which of the CM-probabilistically information-preserving inferences are available to all reasoners once and for all. Instead, we should hope for an account of the *nature* of the relation that explains its features.

This will probably be a challenging task. After all, if there were credal inference it would seem to subsume many of the same kinds of phenomena we overviewed for ordinary inference in Chapter 5. The rationality of credal inference seems to require more than merely being in one credal state after another. There should be an 'accompaniment' to the transition that helps constitute the inference, and that avoids worries about deviant causal chains. We have good reason to anticipate that whatever this accompaniment is, it helps rationalize credal inference, and so would threaten to generate regress worries. After all, what rationalizes the accompaniment? And not only can we not 'jump' to a particular credence in a complex logical or mathematical claim, but ordinary humans should probably arrive at their credences in smaller steps of reasoning. And so on.

What we would hope for is some account of graded representation that naturally gives rise to a cognitive relation addressing these concerns, just as we saw could occur for full attitudes of acceptance in Chapters 4–5. Can such an account be provided? I am unsure. I certainly don't see a natural way to extend the account I've offered to the graded setting. Moreover I am skeptical of the existence of unreduced graded mental states, and so even more so of graded inference, which makes it hard for me to sympathetically pursue the issue further here. I merely want to note that it is ultimately the burden of the defender of graded inference to provide this account.

Second, I have focused here on mental state transitions that bridge two graded states. But what about the relationship between full states (if any remain) and graded states? Information-state semantics like those provided by Yalcin seem to allow that this language can characterize both kinds of states (and even characterize them simultaneously—as when a sentence's truth is sensitive both to a 'world' parameter and an information-state parameter). Are there forms of reasoning that bridge the two kinds of states? If so, do we need to provide *another* kind of foundational account of mental state transitions? If not, would an information-state logic need to 'quarantine' language character-

izing graded from non-graded states, lest it over-generate in the classification of possible rational mental-state transitions?[71]

Third, by taking a relational approach to credal states, we replicate the concerns we saw for MacFarlane in §11.2.2. Recall that once we refine mentality in relational terms, non-trivial questions arise about the relationship between the set of propositions (now probabilistic propositions) a state accepts and the 'information' contained in that state (now a set of CM-probability spaces). We must resolve these issues before we can definitively answer questions about what a logic for our system should look like.

To see how this plays out in the present context, recall that our semantics makes use of an information-state parameter which takes as a value a probability space. A relation like informational consequence tracks information about these values—that is, about probability spaces. So if informational consequence somehow helps to model good inference, we should expect a graded state to eventually be associated with a probability space (or perhaps a set of such spaces), even if this space is not the fundamental description of its structure. But a graded state is now given fundamentally by a set of probabilistic contents. And we have not yet specified a method of getting from a set of probabilistic contents to a unique probability space (or unique set of such spaces). For example, suppose an agent accepts a single proposition: $|\triangle\phi \vee \triangle\neg\phi|$. What information should we say is contained in this agent's credal state? Many different probability spaces support this statement. Some assign a high probability to ϕ. Some assign a high probability to $\neg\phi$. None make ϕ and $\neg\phi$ equiprobable. Symmetry seems to rule out privileging ϕ over $\neg\phi$ or vice versa. But then there is no single probability space left to model the information.

In fact, in the probabilistic setting the problem is pervasive. Suppose an agent accepts only $|.4\phi|$. There are innumerable CM-probability spaces that accept this proposition. Which models the information in the agent's graded state? None seems to be privileged. One option to get out of this problem is by allowing for partially defined probability functions to model the information in a mental state. But this could cause serious disruptions (e.g. should it lead us to redefine informational consequence over spaces with partially defined functions)? And even if not, the issue of symmetry for $|\triangle\phi \vee \triangle\neg\phi|$

[71] An even trickier question: what happens if information-state sensitive language can *itself* become the object of credal assignments? See Goldstein & Santorio (2021) for an exploration of this possibility.

shows the problem goes slightly deeper than this—partially defined functions don't seem of much help in that case.

My sense is that the most promising move here is to say that 'the' probabilistic information encoded in a graded mental state constituted by probabilistic propositions is not given by a single probability space but by a set of them: namely, the set determined by the intersection of any accepted probabilistic contents. In this case, the information contained in a graded mental state is no longer the sort of thing that can occupy the place of an information-state parameter in our compositional semantics (so, for example, the semantics for attitude reports Yalcin sketched above can no longer be the right one). Still, we can use the evaluation of a sentence relative to a single probability space to fruitfully track conditions on good inference by quantifying over that parameter in our consequence relation, just as informational consequence does.

There would then be a tight analogy between the role of the world parameter for truth-conditionally structured states and the role of the information-state parameter for probabilistically structured states. Inference between truth-conditionally structured states requires preservation of correctness relative to a world. Inference between probabilistically structured states requires preservation of correctness relative to a probability space. Just as for truth-conditional inference, where we gain insight into entailment relations by quantifying over the world parameter when assessing relations of correctness preservation at a world, so too for graded inference can gain insight into entailment relations by quantifying over the information-state parameter when assessing relations of correctness preservation at a CM-probability space.

But this discussion brings us to a fourth concern, which could be the most serious one. We've arrived at a framework which treats fundamental attitudinal structure with sets of probabilistic contents, and the information contained within them as a set of measures. And there are serious foundational worries about the interpretation of this kind of framework as one for genuinely graded attitude states. I just noted an analogy between the role of a world parameter and an information-state parameter in our study of entailment relations, including logical ones. But there is a concern that we have more than just an analogy here: we seem to have effectively restructured true graded mentality out of our framework. To understand this, it will be helpful to review some of the ground-level motivations behind the shift to graded mentality.

The basic motivation for accommodating credences that do not reduce to

full acceptance states is the idea that we might not hold a single 'on/off' attitude, but instead lean in degrees toward one of these extremal values. There are various ways of fleshing out the idea, but sometimes matters are framed in terms of *confidence*. Although two agents can count as believing the same proposition, one can be more confident in their belief than another. This seems like a difference that could be reflected in the informational structure of their attitudes. At other times matters are framed in terms of something more like dispositions to believe. Consider two persons neither of whom believes or disbelieves p, but one of whom is on the verge of accepting p—so that only slight evidence in favor of p would tip them over—and another of whom is on the verge of accepting its negation. One could think that this cognitive difference is a representational one, but doesn't trace to attitudes toward content other than p. And belief cannot be the attitude in question, since by hypothesis neither character believes or disbelieves p.

Accordingly, theorists posit graded states—states which take as objects the original content, but where grading resides in the attitude itself. For example, when two people have different degrees of confidence in the same proposition, it is confidence in the *same* content that is at issue. The situation would be analogous to that of other attitudes that uncontroversially bear graded structure like desire, hope, or fear. For example, we might both want a given sports team to win, though you are a die hard fan and I am merely an otherwise indifferent spectator with a trivial sum riding on the outcome of the event. In this case, we prefer one and the same thing: for the team to win. It is the *way* we prefer—the strength with which we prefer—that marks the difference between us.

What is important to know is that as soon as we allow acceptance states that take on probabilistic propositions like $|\triangle\phi \vee \triangle\neg\phi|$ *and* we treat the fundamental structure of the acceptance state in terms of a set of probability measures, we have strayed *very* far from these initial guiding motivations—so far in fact that it is no longer clear they can be motivations for the framework we have ended up in. Allowing acceptance of probabilistic content *without also* structuring mentality fundamentality in terms of a set of measures might have been unproblematic. For this would have been compatible with deriving acceptance of the probabilistic proposition from a more fundamental mentalistic structure given by a single measure, and it is clear that any such measure could in-principle give the structure of a single graded state, borne to the orig-

inal contents. On this reductive approach, believing the probabilistic content consisting of the set of measures that assign high values to rain could just have been the having of sufficiently high credence in rain. The probabilistic component of the content, at the more fundamental level, could reduce to become a component of the structure of the attitude, analogous to the strength of a desire.

But once we take the fundamental structure to be given by sets of measures, this no longer holds. In fact, we are now operating in a framework with (a) on/off states, with no grading or refinement to them (and adding such grading would seem redundant) where (b) the on/off states are borne to 'contents' that have probabilistic elements build into them, that do not reduce away to become part of mental structure at more fundamental levels of description. That is, we have built back into the structure of mentality *all* the key features of 'full' attitude states that the theorist of graded mentality seemed motivated to abandon.

This raises a number of concerns. How can we extend the motivations for graded modality to the new framework? Indeed, if I accept $|\triangle\phi \vee \triangle\neg\phi|$ what is my mental state like? I seem to be nether confident in ϕ (since my state involves some measures where ϕ is improbable), nor confident in $\neg\phi$. And I am also not equally confident in both (in fact, this is the one degree of confidence my state seems to rule out definitively). But then, how do we understand my state in terms of something like degrees of confidence?

Perhaps more worrisome: In what way is the framework actually distinguishable from one where there are fundamentally only full attitudes of acceptance, sometimes borne to truth-conditional contents about probabilistic structure? There is, after all, a kind isomorphism between the two views given (a) and (b) above. Probabilistic contents can sometimes be viewed as constraints on worldly information. From this perspective, it looks suspiciously like all we have done is moved some probabilistic information from a world parameter over to a new parameter and tracked these two pieces of worldly information separately in our semantics.

These are obviously very complex questions about which there is much more to say. What I wanted to flag here is the following interesting trajectory. We have a reasonable good grip on what it is to reason and even infer with full attitudes of acceptance. But there are some important motivations to think we have irreducible graded states of acceptance alongside these. The problem

is that it is highly unclear what it is to reason, let alone infer, with these states. Making sense of this seems not only critical to understanding their rationality, but also to safeguarding their relevance to logic. In trying to flesh out a theory of reasoning it is tempting to simply build back all the structure of ordinary attitudes of acceptance. But at this point it is unclear that we are speaking to our original motivations for the graded framework, or even whether we are giving a different framework at all.

This is obviously not the end of matters, but the beginning. There may be more to say in defense of the framework just arrived at. And there may be other unexplored paths that I have overlooked. Exploring these, however, would take us far afield. We have enough ideas on the table to merit a pause while we try to extract some general lessons.

I think the discussion so far motivates a general moral about the investigation of logics for information-state semantics for probabilistic discourse. This is that we should exhibit *extreme* caution when we try to say what this logic would amount to. The moral of this whole chapter has been that what a logic could be for *any* information-state semantics depends very heavily on the embedding framework for the semantics, as this embedding framework interacts in complex ways with theses about the structure of mentality, which in turn influence how we could understand a logic for the semantics as tracking relations of inference. This is especially important for information-state semantics because the *motivation* of many theorists for adopting such semantics is precisely to allow for the expression of non-truth-conditional mental structure. This general lesson of the chapter applies with added emphasis in the specific context of probabilistic mentality. While in the non-probabilistic setting we can at least see several avenues to pursue (like those we explored in §11.2), in the probabilistic case we are struggling to find even *one* possible interpretation of the formalism consistent with its core motivations.

Perhaps this was to be expected. In Chapter 1, I stressed that the study of inference is still in its infancy. It should then be no surprise that we would be in the dark about how to *extend* of a theory of inference to a complex setting like graded mentality. Still, we should bear in mind that any obstacles in making that extension end up being obstacles to giving sense and purpose to our logical formalism. Accordingly, this points to one more area where a focus on developing and refining our understanding of inference could be absolutely critical for advancing our knowledge of the foundations of logic. Probabilistic

talk could end up being a domain where, without that development and refinement, we end up with no grip whatsoever on what a logic for the relevant discourse could be.

CHAPTER 12

Concluding Remarks

We've covered a lot of ground. Having got deep into the logical weeds with the many applications of Part II it can doubtless be hard to see the forest for the trees. So it is time to step back and review some of general lessons that we can extract from these applications and their interaction with the foundational work of Part I.

In Chapter 1, I suggested that one could read this work in a weaker and a stronger mode. On the weaker mode, one could set aside labels like "logic" and simply view this work as a formal investigation of the properties of good deductive inference which should surely be a worthwhile endeavor whether or not it conforms to any preexisting tradition. On a stronger mode, we can read the result of this investigation as strikingly responsive to a hodgepodge of desiderata for a conception of logic surfacing in core tradition. Let me trace out the grounds for the stronger reading by leading there from the weaker one.

Suppose someone had never heard of logic before, but further that they could distinguish, roughly by ostension, the phenomenon of deductive inference. They set out to investigate the conditions on performing it well that float free of psychologically variable appreciability requirements. What this work has argued is that such a theorist would naturally be led to consider grammatical patterns among sentences that express the contents that undergird good inference. Notably, to the extent they focus on inferential patterns in 'semantically well-behaved domains' like mathematics, they would be specifically be led to the class of classical validities, and perhaps even familiar model-theoretic techniques for capturing them (Ch. 7). They would be open to relaxing the strictures of classical inference rules if they found the possibility of serious semantic defect could infect the language used to characterize the contents figuring in inference. This would lead them to various familiar trivalent logics

depending on how the linguistic facts shook out (Ch. 9), though they might (and would be well within their rights provided they were clear about what they were doing) hold fast to Strawsonian characterizations of consequence in that setting as more suited to their theoretical purposes. If they delved into the thornier territory of inference in the setting of perspectival thought (and they had *de se* exceptionalist sympathies) they would construct a logic LD* almost exactly like Kaplan's logic of demonstratives, though with quite different philosophical backing (Ch. 10). If they were persuaded by linguistic evidence for information-state semantics and the customary interaction of information-state-sensitive language with attitude reports, they would naturally be led to the development of contemporary notions of informational consequence (Ch. 11), though with varying interpretations correlating with whether mentality became restructured with informations-state sensitive contents.

In short, this person who had never heard or seen logic before would be led, by the structure of inference and its relations to language, to *recreate* huge tracts of existing logical practice with identical or near-identical formal structures. What is more, the nature of the phenomenon of deductive inference and its relation to these frameworks would lend them a striking significance. Each of the frameworks would be imbued with *epistemic significance* connected to the role of inference in expanding knowledge and justified belief. And each of the frameworks would be immediately and directly imbued with *normative significance* (Chapter 3). Though this significance would limited by omissions of some instances of entailment relations (like lexical entailments) and the idealization away from psychologically variable facts about appreciability, the normativity in question would otherwise be indefeasible, simple, direct and exceptionless—enshrined in simple principles like my earlier (Good). Moreover, the normativity inherent to the frameworks would bear specifically on reasoning (via inference, as a proper part of such reasoning) in a way that is not shared by other truths, including general truths. The frameworks would always have conceptual ties to necessity—in particular to *metaphysical necessity* (Chs.4, 5)—and through the function of inference we could see that those ties also enabled logic to be an investigation of truths and relations among truths that have the potential to *constrain cognition* (Chs.4).

In short, not only would our imagined theorist be directly led to recreate large and central areas of formal inquiry as logic is currently practiced, but the

nature of the investigation would give the resulting formal investigations the very kinds of implications that have for centuries been held up by myriad logicians and philosophers as capturing what is distinctive and special about logical inquiry.

The tight match between the shape of the formal investigation of deductive inference and that of core logical tradition constitutes powerful evidence for the stronger reading of this work, on which the formal investigation of inference is *among the best rational reconstructions of that core logical tradition available*. It is hard to see how a rival conception of logic could capture significantly more facets of that tradition—there does not seem to be *significantly* more left to capture.

This is *not* to say that there could not be a rival conception of logic that does about as well as the inferential conception, perhaps by being responsive to a slightly different set of historical and traditional demands. For example, perhaps a formal investigation into subject-neutral general truths could play this role as well. I don't take anything I have said in this work to rule out this possibility.

Still, this comes with two qualifications. The first, and less important qualification, is that the inferential conception sets a helpful benchmark for rival conceptions. That is, we would ideally like to see a similar reconstruction, in foundations and applications, in which we set out our investigative task in non-logical terms and gradually rebuild a battery of more or less familiar formal frameworks, and imbue them with the kinds of significance that have at least sometimes been claimed to hold of logical matters.

The second, and more significant qualification, is that *even if* some such conception were to do roughly equal justice to core logical tradition, an inference-based conception of logic would still stand, now merely as one among several equally good reconstructions of that tradition. And this has very important implications for a discipline-wide tendency to explore general skeptical pronouncements about the significance of logical inquiry.

For example, in Chapter 3, we saw Harman argue that logic has no specially normative relevance for reasoning. If that claim is not outright false, it is misleading for only applying to only one of several stipulated conceptions of logical investigation. And given the availability of the inferential conception, Harman's conclusions as applied to some stipulated use of "logic" with no specific ties to reasoning do not obviously have much news value. In Chapter 10,

we saw Russell advance the claim that logical truths are not necessary. Russell did not qualify her remarks by stipulating a particular construal of logicality. Her remarks can at best apply to some reasonable reconstructions of logicality but not others. And even if Russell had made such a stipulative restriction explicit (as, e.g., one could argue Kaplan did), the significance of this claim would be vastly diminished, in light of a rival conception of logic that preserves ties between logicality and necessity even for the domains of discourse involving indexicality. Indexicality does far less work breaking the connections between logicality and necessity than the stipulated conception of what counts as logical.

This is not to mention the implications for individual logical rules. We've seen, e.g., in Chapter 7 that on the inferential conception of logic Ex Falso should be maintained (*even if* it rarely or never constitutes good reasoning, or a good inference). Perhaps in some other context the principle should be rejected. But any such claims should come *qualified* by limiting the rejection of the rule to a formal system that studies something other than the conditions on good inference in the absence of an appreciability requirement.

So much for the importance of the work of this book for logical investigations *generally*. In the next, final remarks, I want to return to focus on the inferential conception of logic specifically to extract a few big picture lessons about logic *so-conceived*. Big picture lessons of this kind can be difficult to discern. After all, one point of this book has been that the foundations of logic on the inferential conception are varied and complex—which creates a strong need to take individual logical principles, let alone logical frameworks, on a case-by-case basis. Even so, some interesting themes have emerged over the course of the past chapters that may have been difficult to see from the individual case studies. Some of these also have further implications for general debates about the nature of 'logic', though I will not explore those connections here, again instead simply focusing on how things look within the inferential conception. Four of these global themes are: the primacy of mentality in dictating the structure of inferential logic; metaphysical necessity as a mark of the logical; the existence of some highly constrained roles of the empirical for logical theorizing; and the existence of clear routes for progress on logical questions.

The first lesson is that on the inferential conception, *mentality comes first*, in the sense that the representational structure of acceptance states dictates the conditions on good inference and so, indirectly, the linguistic relations that are

used to help track it. What this means is that there are no *significant* changes to logic unless one somehow complicates the (standard, truth-conditional) representational structure of acceptance states. (I regard the changes in what vocabulary one counts as 'logical' as insignificant in this context—as changes in which vocabulary counts as logical merely leads to the investigation of broader or narrower sub-classes of a fixed class of good inferences more generally.) Conversely, when the representational structure of mental representation is altered, this gives rise to the possibility of significantly 'new' logics. We saw a particularly striking example of the former of these two complementary lessons in the discussion of semantic defect in Chapter 9. There we saw how the addition of a third truth-value to mark the presence of forms of semantic defect could end up doing *nothing at all* to our consequence relation—not only returning the same inferential patterns, but returning them as the result of a consequence relation that, on reduction, turned out to be *conceptually* equivalent to the classical one. This was precisely because the forms of semantic defect posited did not directly impinge on mental structure. A similar lesson arose in the discussion of information-state logics for modals and conditionals in Chapter 11. There we saw how the most perspicuous logic for the setting given by Yalcin's construal of the operations of conditionals and modals as characterizing total mental content precluded those operations from helping to characterize good inference proper, thereby leading to classical logic as the most perspicuous inferential logic in that setting. As it happens, we also saw the complementary lesson in that chapter: that adding structure to mentality forces us to reconceive our understanding of inference in a way that pushes down into logic. This occurred to a lesser extent for MacFarlane's embedding framework for conditionals and epistemic modals, and to a much greater extent for investigations of probabilistic mentality. Indeed, this last framework so radically restructures mental representation that it was unclear how to recover an inference based logic at all.

The second lesson is that metaphysical necessity is a mark of the logical, but that we can get systematic illusions to the contrary by failure to attend to the multifarious relations that can exist between language and thought. We in fact saw numerous variations on this lesson in Part II. In Chapter 8, we saw how the appearance of contingent logical truths could be generated in a modal logic by introducing a mismatch between the truth-conditional profiles of the assertoric contents of sentences and the profiles evaluated by a necessity oper-

ator. All that this mismatch showed was that the modal operator tracked the necessity (even the metaphysical necessity) of something other than the mental contents that figure in good deduction. There was accordingly no tension in claiming that ϕ is a logical truth (accordingly) expressing a metaphysically necessary content, and that $\Box\phi$ is false (at actuality, in a model) for a *kind* of metaphysical necessity modal $\Box$. In Chapter 10, we saw that we could get contingent logical truths for a distinctively Kaplanian conception of logic, but that there were strong grounds to think this conception of logic was divorced from good inference. Reintroducing such a conception tied logical truth back to a generalization of metaphysical necessity in the context of *de se* cognition, and restored the significance of metaphysical necessity even for context-sensitive language of both perspectival and non-perspectival sorts. Finally, in Chapter 11, we saw Bledin argue that logics tracking good inference for information-state-sensitive language would use informational consequence, which did not track necessary preservation of truth. But once the mechanics of information-state sensitive language were attended to—in particular, insofar as they are used to characterize total mental structure—we saw that information consequence in fact gained its utility precisely by being the appropriate technique for generalizing methods of tracking metaphysically necessary truth-conditional structure. In each of these three cases the same underlying mistake lurks behind the illusion of logicality without necessary truth-preservation: that of using language to track something other than the attitudinal contents that underlie inferential relations. Instead one might track other profiles of truth-conditions (as in the modal case), or items that are not contents at all (as in the indexical case), or one may track global representational features of mental states (as in information-state semantics). This general lesson about metaphysical necessity here dovetails with the first on mentality: without a change in the structure of representational content that underlies good inference, a change in how one uses language can at most shift the target of inquiry away from good inference—it cannot change what makes an inference good. Accordingly it cannot effect a fundamental change in logic, insofar as this is an investigation into good inference. At best it could constitute a new way to use the term "logic".

A third lesson concerns the relation of empirical investigations to logical inquiry. There has been noteworthy debate about the role that empirical information has to play in logic, and whether (for example) logic in the end is

'continuous with the sciences.'[1] I suspect some of this debate is muddied by unclarity about what falls under the heading of logic, and I certainly won't be able to do full justice to the scope of these debates in a few remarks. I only wish to note how empirical considerations *have* come to play a role on the inferential conception. This occurs by two paths of influence: a more minor path through linguistics and the philosophy of language, and a more major path through the philosophy of mind. The influence through linguistics arises because we study deductive inference through language. Because we want to understand the *actual* inferences we perform, we have no other way, since it is only through the actual use of language that we have any grip on the nature of the actual thought contents that figure as the starting- and end-points of our deductions. Of course, we may use stipulated 'formal' languages to simplify the exploration of these thought contents. But eventually the truth-conditional contents expressed in these formal languages must be brought back into contact with our actual inferential practices if they are to have any significance.

We saw one aspect of this impingement of the empirical on the logical in our discussion of semantic defect from Ch. 9. Now, there are important conceptual questions about the very possibility of certain forms of semantic defect (e.g., whether assertoric contents could even in principle bear 'contingent-defect-at-a-world') which one could argue are not, strictly speaking, empirical matters. But granted the possibility of various forms of defect there would certainly remain empirical questions about where in language such defect arises, how it arises, and especially how it compositionally projects. The outcome of these empirical questions could have a large role to play in logic, revealing that simple logics maintain their simplicity only by stipulating away linguistic complexity. But they could equally swing in the other direction, and reveal that much larger tracts of natural language (and so ordinary inference) fall within the domain of the 'semantically well-behaved.'

This kind of influence of the empirical on the logical is constrained by the earlier observation lesson 'mentality comes first.' The changes in logic we obtain depending on how the linguistic facts shake out do not (on their own) change anything of fundamental significance for logic—they do not change the general nature and form of deductive inference itself, just how we track it. However, a second influence of empirical theorizing has precisely the power to

[1] I am thinking here of debates between exceptionalists and anti-exceptionalists about logic. See, e.g., Ferrari et al. (2023) and the citations therein.

shift the foundations of logic itself in this way. We saw one instance of this in Chapter 8 in the discussion of puzzles for inference created by the presence of rigidifying descriptions. One version of these puzzles challenges fundamental assumptions about the nature of good inference itself—e.g. the principle that the goodness of an inference depends merely on the contents that the inference mediates between (what I there termed "UNIFORMITY"). We saw that a wide range of empirical data on conditionals, counterfactual discourse, and negative existentials ended up bearing on this issue. The outcome of various empirical questions had the power to reshape our conception of inference itself. An even more striking instance of this influence of the empirical came at the end of Chapter 11, where we saw that the motivations (doubtless many of them empirical) for structuring mental representation with *graded* attitude states ended up disrupting a conception of inference so dramatically that we could not obviously recover any inferential conception of logic at all.

In short, if we stipulate that logic studies aspects of cognition via their expression through language, empirical findings on language and cognition could obviously have the potential to alter the structure or conception of logic that results, sometimes by influencing the best ways to track good deduction, sometimes by altering our conception of good deduction itself.

Though there is clearly room for empirical results to play these roles on the view I've been developing, it is worth emphasizing that all these forms of influence arise because of the explananda we have stipulated that we wish our theory to be responsive to, and that these explananda constrain the role that empirical matters could influence. Sometimes philosophers have suggested that we can arbitrate between logical theories on the basis of tie-breaking metrics from the sciences like overall simplicity of theory—which might favor 'simpler' logics like classical logic. We have not seem room for this kind of influence on the inferential conception, nor is there any reason to think these modes of arbitration would have much impact. This is because the explananda to which logics are by stipulation responsive are so rich, leaving little room for substantially different competing theories to handle it equally well. We could never favor classical logic for its simplicity in this setting any more than we could favor Newtonian Mechanics over General Relativity on grounds that the former is simpler. It's just clear in the physical setting that responsiveness to the data takes center stage. The same would be true of any competition between classical and non-classical logics in the inferential setting.

This is connected with the final lesson of the foregoing chapters I want to highlight: the richness of non-logical foundations for logical inquiry on the inferential conception. Throughout this book I have flagged innumerable loose ends—places where my argumentation gave out and would have to pick up in other places to be completed. We left open multiple questions in Chapter 7 about whether and how the 'semantically well-behaved' features of a domain like mathematics could generalize to other areas of discourse. We left open questions about the nature of assertoric content in Chapter 8 that influence the nature of good deduction. In Chapter 10, we tentatively explored what followed from *de se* exceptionalist views, noting that it remained highly controversial whether exceptionalism held in this domain, and if it did what exact shape it would take. In Chapter 11, I left open whether linguistic data motivated information-state semantics, what shape they would take, how they would related to mentality (e.g., as Yalcin suggested, or as MacFarlane suggested, or yet in some further way), and also left open the issue of whether and how mentality should be fundamentally graded with probabilistic structure.

To say that leaves much unresolved is probably an understatement. The goal of these chapters was not to settle logical issues definitively, but to provide productive non-logical routes to settle them. In all the above examples, logical frameworks and principles are at stake. But, critically, the adjudication of those frameworks and principles does not itself rest on logical debates. This, to me, is one of the most attractive features of the inferential conception of logic: that it provides so many productive routes of inquiry to settle broadly logical questions. In this way, the raft of threads I've left dangling is the inevitable cost of giving diverse foundations to logic. Tying logic to substantive issues in metaphysics, philosophy of language, linguistics, and philosophy of mind is bound to embroil the logician in issues that cannot be resolved in the course of a single book. This is just the price for locating rich non-logical grounds for logicality. And it is a price I think we should be willing to pay. The bought potential for definitive progress in the foundations of logic is too enticing, and I fail to see how that progress can be paid for in any other way.

Part III

Appendices

Appendix A

Experimental Set-Up

North American consultants were recruited through Amazon Mechanical Turk in the spring of 2021 for a series of three separate online surveys which respectively requested those consultants to rate how easy or difficult they found supposing, imagining, or visualizing an array of figures meeting certain specified conditions. The number of consultants initially recruited for each task were: 144 (supposition task), 128 (imagination task), and 128 (visualization task).

The text of the prompt given to consultants for the supposition task was as follows (prompts for the the visualization and imagination tasks were similar but with "visualization"/"imagination" talk substituted for "supposition" talk).

> This questionnaire begins by presenting several English sentences, some of which may be quite complex and some of which may sound odd. Your task is to suppose to the best of your ability what the sentence reports, and to report on how easy or how difficult you found supposing it. Please indicate how easy or hard it was to suppose what the sentence reports by selecting one of the buttons below the sentence, where 1 is "I could not suppose the scenario" and 7 is "it was very easy to suppose the scenario".
>
> Tasks are to be interpreted against the following background: there are five figures in a row named A, B, C, D, and E. They are arranged in a straight line with Figure A furthest to the left, followed by B, C, and D, and finally with E furthest to the right. So A is next to only B, B is next to only A and C, and so on. Each

> of these five figures is a square, a circle, or a triangle. There are no other types of figures.
>
> Each question asks you to suppose several things at the same time of a single arrangement of figures. So if a question reads:
>
> *Suppose that Figure A is a triangle and that Figure B is a square.*
>
> You are being asked to suppose that there is a single arrangement of five figures, the first of which is a triangle, and the second of which is a square.
>
> Here is an example:
>
> *Suppose that figure A is a circle.*
>
> 10 20 30 40 50 60 70
>
> In the above example, you will probably have little difficulty supposing what is asked. If so, please select a button on the right, such as 7.
>
> Some sentences may be harder for you to suppose. If so, you would select a button further to the left accordingly.

This formulation strictly speaking tests speakers for supposition/imagination/visualization of contents more complex than is given by the example sentences (since they are also asked to suppose/etc. the background facts about figures, numbers, shapes, etc.).

Because of the complexity of the task, consultants were asked to complete a brief questionnaire to test their understanding of the directions. This test consisted of three questions given in the following prompt.

> Before you begin, please answer the following simple questions to verify your understanding of the task.
>
> How many figures are you to suppose are in a given row of figures?

3 ○ 4 ○ 5 ○ 6 ○

Are you ever to suppose there are pentagons among the figures?

yes ○ no ○

If you find that you cannot suppose what you are asked to suppose, what number should you use to indicate this?

1 ○ 7 ○

If a consultant was not able to answer all the above questions correctly ("5", "no", "1"), their further responses were ignored in the results tabulated below and discussed in Chapter 4. The number of participants who answered the verifying questions were: 77 (supposition task), 63 (imagination task), and 58 (visualization task).

Consultants were then presented with a series of four sentences randomized along two dimensions: logico-syntactic complexity, and possibility/impossibility. As noted in Chapter 4, logico-syntatic complexity is merely a useful surrogate in this context for the representational complexity that is of interest. There were four 'degrees' of complexity ascertained to be in ascending order of complexity in intuitive terms, with no presumption that changes in complexity were 'evenly spaced.' Each consultant received at least one sentence from each degree of complexity. Otherwise sentences were randomized (both in order of complexity, and in whether they were possible or impossible).

Here is a list of the sentences that were drawn from in the supposition task (the same sentences were used, suitably modified, for the other tasks).

Complexity 1

possible: Suppose that Figure A is a circle or a triangle.

impossible: Suppose that Figure A both is and is not a circle.

Complexity 2

possible: Suppose that Figure C is a circle or a triangle, and Figure D is neither a circle nor a triangle.

impossible: Suppose that Figure C is a circle or a triangle, and Figure C is neither a circle nor a triangle.

Complexity 3

possible: Suppose that Figure C is a square, that every triangle is next to at least one circle, and that the only figures next to squares are triangles.

impossible: Suppose that Figure D is a square, that every triangle is next to at least one circle, and that the only figures next to squares are triangles.

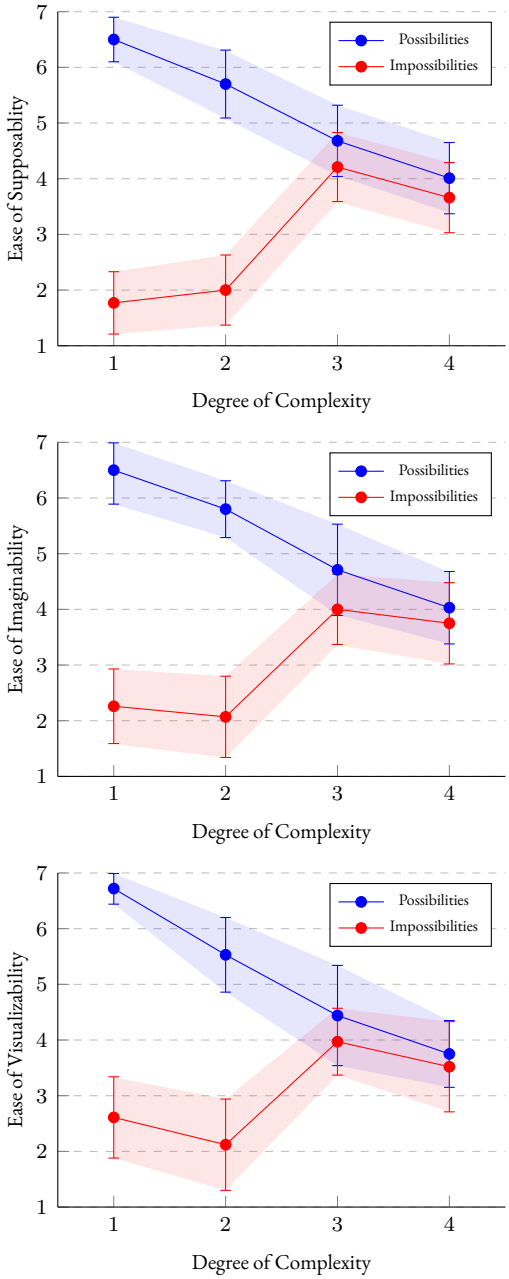
Complexity 4

possible: Suppose that Figure B is next to at least one circle, that every figure next to a circle is also next to a square, that Figure D is next to a triangle, and that Figure D is not a circle.

impossible: Suppose that Figure B is next to at least one circle, that every figure next to a circle is also next to a square, that Figure D is next to a triangle, and that Figure A is not a circle.

Results summarized in Figure 4.1 of Chapter 4, repeated here, show the mean values of responses, alongside 95% ($\alpha = .05$) confidence intervals (for unknown population standard deviation).

FIGURE 4.1 Ease of supposability, imaginability, and visualizability plotted against complexity of content



Appendix B

A Kreiselian 'Squeeze' for Unrestricted Quantification

Why think that model-theoretic validity extensionally tracks a 'true' conception of validity—one that respects the possibility of unrestricted quantification? After all, models inherently restrict quantification to set-sized domains, and unrestricted quantification is not so-limited. Why think truth in all models gives us information about truth in quantificational contexts witnessed in no model?

An influential approach to this problem is articulated in Kreisel (1967). Kreisel suggests we can make headway by taking validity as a primitive, and leveraging intuitions about the relationship between this primitive concept and those of derivability and truth-in-all-models.

Consider any system of derivation for first-order logic that is sound and complete with respect to model-theoretic validity. Let $S_{\text{derivable}}$ be the set of sentences derivable in the deductive system, $S_{\text{mt-valid}}$ be the set of sentences true in all first-order models (which we presume to have set-sized domains), and S_{valid} be the set of sentences that express 'intuitively valid' truths—in particular, valid even given the possibility of unrestricted quantification.

Kreisel suggests that the axioms of our derivation system are intuitively valid, and the rules intuitive-validity-preserving. If so we have the following containment.

(i) $S_{\text{derivable}} \subseteq S_{\text{valid}}$

What is more, if we could find a model in which a sentence came out false, that would suffice to show that the sentence could not be an intuitive validity. That gives us the following containment.

(ii) $S_{\text{valid}} \subseteq S_{\text{mt-valid}}$

But since our proof system is complete with respect to the model-theory, we have the following containment.

(iii) $S_{\text{mt-valid}} \subseteq S_{\text{derivable}}$

From all three containments we obtain an equivalence.

(C) $S_{\text{valid}} = S_{\text{mt-valid}}$

I of course have recommended against taking validity as an undefined primitive in the context of an exploration of deductive inference. Complicating matters, I have suggested that validity is inherently linguistically relativized—that is, relativized to a particular combination of syntactic and semantic properties. In particular, in Chapter 7 I introduced a special batch of linguistic properties under the heading of 'modalized first-order form,' and suggested that model-theoretic validity could be thought of as tracking this 'true' form of validity—MFOF-validity. Still, even given these departures from Kreisel's starting point, the general form of his argument can carry over to this new setting in a straightforward way.

It was stipulated to be part of a sentence's modalized first-order form that its quantifiers be restricted to a set-sized domain. But we can obviously relax that assumption. We can characterize an *expanded modalized interpretation* as a modalized interpretation modified to allow for (but not require) unrestricted quantification provided that notion makes adequate sense. We can then define the *expanded modalized first-order form of a first-order sentence* ϕ to be the set of syntactic and base semantic properties shared by ϕ on all of its expanded modalized interpretations. Say that a first-order sentence is *MFOF$^+$-valid* if, necessarily, on any interpretation of ϕ that gives ϕ expanded modalized first-order form, ϕ expresses a necessary truth. Truth in all models (with set-sized domains) tracks MFOF-validity. Does it *also* track MFOF$^+$-validity?

Let $S_{\text{MFOF}^+\text{-valid}}$ be the set of first-order MFOF$^+$-valid sentences. Then we can give informal (if not 'intuitive') justifications for taking the first containment to hold: the properties of extended modalized first order form seem to explain the expression of necessities by axioms of first-order derivational systems, and to explain how the rules preserve the expression of necessities in virtue of those properties.

(i′) $S_{\text{derivable}} \subseteq S_{\text{MFOF}^+\text{-valid}}$

The analog of the second containment does not need to hold on intuitive or informal grounds. Any model with a set-sized domain in which a first order sentence is false would be one that *witnessed* the failure of MFOF$^+$-validity. For any such model could trivially be extended to a (non-extended) modalized interpretation falsifying ϕ (at actuality). This sentence, on this modalized interpretation, would possess modalized first-order form. And any interpreted sentence with modalized-first order form has extended modalized first-order form by definition (since the latter essentially only introduces disjunctions to select properties of modalized first-order form). So (ii′) is actually true by definition.[1]

(ii′) $S_{\text{MFOF}^+\text{-valid}} \subseteq S_{\text{mt-valid}}$

Our completeness theorem in (iii) above remains relevant, and unchanged, and from (i′), (ii′), and (iii), (C′) follows.

(C) $S_{\text{MFOF}^+\text{-valid}} = S_{\text{mt-valid}}$

On one way of looking at things, the Kreiselian argument is strengthened for dropping any appeal to intuition at the second step—something facilitated by having an independent reduction of logical consequence. On another way of looking at things, the argument is slightly trivialized, and less informative, precisely for losing that intuitive appeal. There is an impression that more of the problem has been 'defined away.' I think each way of looking at things has an element of the truth, and I certainly would not shy away from the second perspective insofar as it accords with my view, discussed in Chapter 7, that much of the work of classical validity is achieved through semantic stipulation.

[1] This is straightforwardly true by definition if we model the 'in virtue of' relation as I have been by necessitation. But even if this is replaced by a more suitable explanatory relation, it will remain straightforwardly true as long as it is necessary condition of the form of explanation that it necessitates, which I presume in this context.

Appendix C

A Concern for Kaplan's Logic of True Demonstratives

In "Afterthoughts," Kaplan engages with a problem for understanding the behavior of true demonstratives. He had always recognized that a demonstrative needed to be 'completed' by a demonstration, construed as a feature of a speech act context, to acquire something like a character. But only in "Afterthoughts" does he explore adjustments needed to accommodate the absence of demonstrations from contexts of utterance. The formal adjustments he does make, I will argue, lead to some striking logical aberrations (e.g. the possibility that a valid conjunction does not have valid conjuncts).

Throughout Kaplan's career, he treats demonstrations as parts of speech act contexts. Early on he treats them as something like public acts of pointing,[1] and later as more 'internal' directing intentions.[2] In Kaplan's late formalism, we model demonstratives with indices to track which demonstrations are 'sought' by the demonstrative to complete their meanings. This is not only because there are intrasentential shifts in which demonstrations fix the contribution of a demonstrative type, but because it is *part of the meaning* of the demonstrative that it requires a directing intention: "the meaning of a demonstrative requires that each syntactic occurrence be associated with a directing intention."[3] (This gives Kaplan's grounds for indexing demonstratives, but not temporal indexicals like "now" or "today".) The result is that "within the formal syntax we must have not one demonstrative "you", but a sequence of demonstratives, "you$_1$", "you$_2$", etc."[4]

[1] Kaplan (1989b, 489–91).
[2] Kaplan (1989a, 582–4).
[3] Kaplan (1989a, 587)
[4] Kaplan (1989a, 587)

In "Demonstratives", Kaplan adds to formal contexts sequences of demonstrata which represent the objects picked out by demonstrations in that context. A demonstrative like "that$_i$" evaluated at a context directly refers to whatever demonstratum occupies the ith place in the sequence of demonstrata. But in "Afterthoughts," Kaplan acknowledges that this is an idealization, since it papers over the ways in which demonstratives can suffer from special kinds of defect that other indexicals cannot.

> [To accommodate demonstratives like "you"] [t]he idea is that the context simply be enriched by adding a new feature, which we might call the *addressee*. But suppose there is no addressee. Suppose the agent intends no one, e.g., Thomas Jefferson, dining alone, or surrounded by friends but not *addressing* any of them. Or, suppose the agent is hallucinatory and, though addressing 'someone', no one is there. The problem is that there is no *natural* addressee in such contexts, and thus no natural feature to provide within a formal semantics.
>
> …
>
> There are really two problems here, calling for separate solutions. The first is the case of the absent intention. In this case one would want to mark the context as *inappropriate* for an occurrence of "you", and redefine validity as truth-in-all-*appropriate*-possible-contexts. The second is the case of the hallucinatory agent. Here the context seems appropriate enough, the agent is making no *linguistic* mistake in using "you". But the occurrence should be given a 'null' referent.
>
> (Kaplan, 1989a, 585–6)

Accordingly Kaplan changes formal contexts by adding two kinds of elements to sequences of demonstrata: null elements (to mark the presence of demonstrations that fail to demonstrate), and inappropriateness markers (to mark the absence of an appropriate demonstration): "within the formal semantics the context must supply not a single addressee, but a sequence of addressees, some of which may be 'null' and all but a finite number of which would presumably be marked *inappropriate*."[5]

[5] Kaplan (1989a, 587).

Note that in the quotation above, Kaplan suggests we must *redefine* validity. It is no longer truth in all proper contexts, but truth-in-all-*appropriate*-proper-contexts. This seems advisable. If we retain the old definition of validity, then the logic of demonstratives will become vacuous. For any demonstrative there will always be at least one context where it is inappropriate, leading to a severe form of linguistic defect. This, presumably, would preclude its contributing to the expression of truths, at the very least for atomic sentences. Thus a sentence like "$\text{that}_1 = \text{that}_1$" would not come out as valid. By contrast, this sentence is valid if validity is assessed relative to appropriate contexts (assuming, as holds in Kaplan's formalism, that an identity between 'nulls' is a truth).

But note that appropriateness is *sentence-relative*. The appropriate contexts for "EXIST(I)" are all proper contexts. It is not possible for this sentence to be inappropriate at a context. But the appropriate contexts for "$\text{EXIST}(\text{you}_1)$" or "$\text{EXIST}(\text{that}_1)$" are a subset of these: e.g., those contexts where the agent of the context tries to address someone, or directs their attention in a demonstration. But this means that *validity* is also a sentence-relative notion. And this leads validity to be sensitive to features of speech act contexts in ways that Kaplan sought to avoid.

We can see a simple manifestation of the problem if we allow ourselves to treat some additional expressions as logical. Let "demonstrate" be a monadic predicate that is true of an object at a time in a world just in case that object is attempting to demonstrate something at that time, in that world. Allow that we can provisionally treat this as a logical predicate. Then because there are proper contexts where agents do not demonstrate anything, we have.

$$\not\models \text{demonstrate}(I)$$

Next consider a sentence like "that_1=that_1". The only appropriate contexts for the sentence are ones where there is no inappropriateness marker in the first slot in the context's sequence of demonstrata. The rest of Kaplan's rules ensure this is valid (as Kaplan clearly intended).

$$\models \text{that}_1 = \text{that}_1$$

But consider now "that_1=$\text{that}_1 \wedge \text{demonstrate}(I)$". The appropriate contexts for this sentence are again ones where there is no inappropriateness marker in

the first slot in the context's sequence of demonstrata. But in all such contexts, the agent of the context is demonstrating. So we have:[6]

$$\models \text{that}_1 = \text{that}_1 \wedge \text{demonstrate}(I)$$

In this way, we obtain a valid conjunction whose second conjunct is not valid. The problem is that the contexts used to determine the validity of the conjunction is a proper subset of those used to determine the validity of that second conjunct.

How does this problem arise? We can see the issue more clearly by focusing on theses about characters, which are the effectively the bearers of validity for Kaplan. I will speak of such characters as composing using conjunction in the obvious way. (That is, for characters χ and χ': χ-and-χ' expresses a truth at a context just in case each of χ and χ' express truths at that context.)

Then we can argue as follows.

(1) For every character χ, there is a range of contexts C_χ such that χ is valid iff χ expresses a truth at all contexts in C_χ.[7]

(2) For any two characters χ and χ', $C_{\chi\text{-and-}\chi'} = C_\chi \cap C_{\chi'}$.

(3) The English "that is that" sometimes bears a character χ_1 that is valid.

(4) "that is that" never bears a character expressing a truth at contexts where the speaker of the context is not demonstrating.

(5) The English "I am demonstrating" always bears a character χ_2 which expresses a truth at every context where the speaker of the context is demonstrating.

These entail:

(C) χ_1-and-χ_2 is valid.

By (1)–(3), $C_{\chi_1\text{-and-}\chi_2} \subseteq C_{\chi_1}$. Then by (4) and (5) χ_2 expresses a truth in every context in $C_{\chi_1\text{-and-}\chi_2}$. By the rules for conjunction χ_1-and-χ_2 expresses a truth at every context in $C_{\chi_1\text{-and-}\chi_2}$.

[6] Cf. Salmon (2002, n.39).

[7] (1) is only true for sentences containing only logical vocabulary. But we are only considering such sentences here.

Rejecting (1) would require a wholesale rejection of Kaplan's conception of validity. (3) cannot be rejected without effectively rendering the logic of true demonstratives vacuous. (5) is also not subject to debate. We can tinker with (2) to avoid this particular problem. But doing so usually generates other troubles. E.g., if we take instead $C_{\chi\text{-and-}\chi'} = C_\chi \cup C_{\chi'}$, then we will not get a validity corresponding to Kaplan's "that$_1$ is that$_1$ and that$_2$ is that$_2$" (as contexts appropriate for the first conjunct may be inappropriate for the second, and vice-versa). And, at any rate, these alterations are *ad hoc* in the context of Kaplan's motivating framework.

So the real issue is (4). It is worth noting that an alternative treatment of demonstratives Kaplan considered in "Demonstratives" avoids the problems we are encountering here, as a formal matter. On that treatment, the 'logically true' readings of the English "that is that" are captured by the validity of formal sentences of the following form, where "*dthat*" is a rigidifying operator and "δ_i" is some description contributed by a demonstration in a given context.

$$dthat[\delta_i] = dthat[\delta_i]$$

The only residual indexicals in δ_i are those which get a value at a context independently of any demonstrations. Accordingly, every element in the sentence always bears content at every proper context (including those without demonstrations).

Note also that we can distinguish between linguistic failure for a demonstrative and 'mere' reference failure on this view. Linguistic failure would simply correspond to a sentence where a *dthat*-operator is not completed by a description—it is a kind of syntactic failure. And reference failure in a context would involve completion of the *dthat*-operator by a description that has no referent when evaluated at that context.

Does this mean Kaplan can get out of his troubles? Matters are not quite so straightforward. Kaplan moved away from the operator formalism for demonstratives because it misleadingly presents the contributions of demonstrations in syntactic terms, and also obscures the fact that demonstratives are devices of direct reference.[8] But these are not the only worries for the formulation.

Suppose I use an English sentence like "you are F" in a context c_1 where I address a single agent (I direct my attention to them, point at them, etc.). Call the character of my "you", 'completed' by the demonstration, ξ. Consider a

[8] See especially the discussion at Kaplan (1989a, 579–82).

further context c_2. c_2 involves a completely different speech setting, with a different agent at a different time and world, surrounded by different objects (with a different spatial distribution and visual presentation) and no possible ordinary addressee. The agent demonstrates nothing (e.g. does not point, or direct their attention at anything, etc.). Nonetheless the agent of c_2 utters "you are F".

What is the content of the character of ξ at c_2? On an operator view it is either some demonstratum, or the value corresponding to a failed demonstration (as when one points at a hallucinated object). But, to the extent the question I've asked makes any sense, it feels like this is getting something wrong. It feels as hard to evaluate the ξ at c_2 as it does to evaluate the *speaker-of-*c_2's utterance of "you are F" at c_2. The latter utterance is an abject linguistic failure (as could hold within either of Kaplan's formalisms). The attempt to evaluate the character of the original utterance from c_1 at c_2 feels like this as well. This is something that Kaplan's later system gets right. As Kaplan would model things, the character of "you are F", as used at c_1, 'seeks' a mode of address at c_2 and finds none there. Accordingly, the character will be assigned the same defect that belongs to the idle use of "you are F" by the speaker of c_2. But note that it is precisely this strategy for treating demonstratives which underlies a claim like (4) as used in the deduction above. So while we can get out the problem by rejecting (4), that rejected claim would still have an important intuitive basis. We need more of a justification to reject (4) than that it helps us regularize the behavior of our preferred definition of validity.

There is another diagnosis of our problems here that draws on the morals of Chapter 10, where I argued there that Kaplan's logic conflates aspects of perspectival thought and linguistic context-sensitivity.

Demonstrations are parts of speech act contexts. And it is clear that linguistic demonstratives have no interesting semantics (e.g., that could be the basis for evaluations of validity) independently of some information from speech act contexts given by such demonstrations. For this reason, Kaplan is forced to feed his logic *some* information about speech act contexts to even begin to consider questions about a logic for true demonstratives. English sentences containing demonstratives like "that" or "you" are not, on their own and independently of demonstrations within speech act contexts, candidates to bear logical properties.

As I argued in §10.5, this kind of maneuver of feeding contextual informa-

tion into a logic actually happens all the time, surreptitiously, when we investigate the logic of context-sensitive quantifiers or modals. But, as I also stressed there, the logic for these expressions checks for something like the metaphysical necessity of the *contents* of what is expressed, given the contextual information. Because Kaplan treats validity as a property of character which is not *based* in properties of content, demonstratives put him in a bind. He must resolve some information about context to arrive at an object that is a plausible candidate for logical evaluation. But once he does this, he must somehow reconstruct validity by looking back to the space of contexts from which we have provisionally thrown away elements. This leads directly to the problems I just brought out.

On the view I favor, we should split questions about demonstratives into two questions, one about the logic for directed perspectival thought about objects (*independent* of language), another about the logic of purely linguistic context-sensitive terms. Sketched *very* roughly, these look as follows.

For a perspectival logic, demonstratives correspond to perspectival thoughts about objects, or more specifically, thoughts which characterize some 'component' of a perspective at which an object may lie. (Intuitively this will be a 'focal' point of the perspective, though I will not assume the 'focus' itself is part of the perspective, or at least not part of the relevant perspectival content). I will assume for now that each perspective has the same possible components of this kind (e.g. locations in a visual field, say).[9] Consider a demonstrative thought directed thrice to the same object in a visual field, to the effect that if the object exists, it is itself. This thought corresponds to a set of center-neutral worlds at which either (i) nothing lies at the relevant component of the perspective (in which case, we can suppose the perspectival thought as intended is true), or (ii) an object does exist at the relevant component of the perspective, and of course will be identical to itself (in which case the perspectival thought is also true). In this case, the *de se* thought in question can be inferred from no premises. And a 'logic' for perspectival thought could capture this, provided it made enough semantic properties of the sentence modeling the thought 'logical' (e.g. by treating as logical the fact that one component of the perspective is picked out three times over, perhaps by using co-indexed demonstratives, etc.). Note that when we treat the thought as logical, we do so precisely *because it is*

[9] It's not clear this is a safe assumption, since we need these components to exist even if there is no agent in a center-neutral piece of information. If the existence of components is not guaranteed, the claims about logicality to follow would fail, and could only be restored by conditionalizing on the existence of the requisite perspectival components.

necessary—given the generalization of metaphysical necessity for *de se* information. That is, no matter which linguistic properties we regard as logical, we ascertain logicality by quantifying over the *full* range of 'contexts,' or what I called 'center-neutral worlds.' What happens to a thought corresponding to a sentence like "if that$_1$ exists, then that$_1$ is that$_1$ and I am demonstrating"? This will not be a logical truth. This is precisely because as we quantify over the full range of center-neutral worlds, we quantify over some where (e.g.) an object exists at the relevant component of a perspective, while there is no agent of the associated center-neutral world (and so no agent demonstrating).

A logic for demonstratives treated as purely *linguistic*, context-sensitive devices of reference could look quite different. Here, unless we restrict the range of contexts we consider, the logic of demonstratives will become vacuous for the reasons that Kaplan notes. That, of course, does not mean that we cannot restrict that range contexts. We could (roughly as Kaplan suggests) restrict our attention to 'appropriate' contexts for the use of certain demonstratives, if we find that illuminating. In this case, "if that$_1$ exists, that$_1$ is that$_1$" could be able to express a validity, in roughly the way that universal instantiation for a context-sensitive universal quantifier can be valid. Relative to contextual information that there is a speaker who has attempted to demonstrate an object (and would have demonstrated the same object three times if they designated any), it would be valid just because any given object is necessarily self-identical. But note that it is the metaphysical necessity of what the sentence expresses (relative to various 'full' contextual resolutions) that would safeguard its logicality. And we again quantify over the *full* range of metaphysical possibilities to assess validity. Because of this "if that$_1$ exists, that$_1$ is that$_1$ and I am demonstrating" will not come out valid in this setting either. Even though we may restrict the range of contexts we consider to evaluate this sentence's validity, that restriction does not come with a corresponding restriction over the space of metaphysically possible worlds used to evaluate validity. At some metaphysical possibilities the speaker of the context will not exist.

Note that a logic for perspectival thought assesses validity by quantifying over formal objects like Kaplan's contexts. But the philosophical basis for the logic *absolutely forbids* restricting the range of contexts, for any purpose, in assessing properties like validity. By contrast, the logic for linguistic context-sensitivity allows us to restrict the range of contexts we consider whenever we find this theoretically convenient or illuminating. But the philosophical

grounds for allowing this are precisely that in restricting contexts, we are not restricting to any extent the range of values used to check for validity.

Once we understand this, the problem for Kaplan becomes more apparent. There is no well-grounded conception of logic that *both* incorporates some information from 'contexts' *and* continues to assess logical properties by quantifying over them.[10] But once we conflate perspectival thought and context-sensitivity in the way Kaplan's framing of logic does, we can feel pressured to do both of these things. Focusing on demonstratives as mere bits of language, which need information from context 'merely to disambiguate,' we feel compelled to use information from speech act contexts prior to asking questions about logic. But because we motivated the logic of our language leaning on its connections with perspectivally guaranteed truths, we feel compelled to somehow continue quantifying over the contexts we have restricted to assess for validity (what other option do we have?). The results are, predictably, problematic.

[10] This also helps us understand the great awkwardness of even asking questions, like those I was forced to ask above, about how to evaluate the character associated with a linguistic demonstrative at contexts incompatible with anything like the demonstration that actually completes it. A logic that forces us to confront such questions has already arguably gone astray.

Bibliography

Adler, Jonathan E. 1999. "The Ethics of Belief: Off the Wrong Track." *Midwest Studies in Philosophy*, vol. 23 (1): 267–285. doi:10.1111/1475-4975.00014.

—. 2002. "Akratic Believing?" *Philosophical Studies*, vol. 110 (1): 1–27. doi:10.1023/A:1019823330245.

Anderson, R. Lanier. 2005. "Neo-Kantianism and the Roots of Anti-Psychologism." *British Journal for the History of Philosophy*, vol. 13 (2): 287–323. doi:10.1080/09608780500069319.

Aquinas, Thomas. 1981. *Summa Theologica*. Westminster, Md.

Aristotle. 1991. *The Complete Works of Aristotle Vol. II*. Princeton University Press.

Armstrong, D. M. 1989. *A Combinatorial Theory of Possibility*. Cambridge University Press.

—. 1997. *A World of States of Affairs*. Cambridge University Press.

Armstrong, David. 1968. *A Materialist Theory of Mind*. Routledge, New York.

Barwise, Jon & Robin Cooper. 1981. "Generalized Quantifiers and Natural Language." *Linguistics and Philosophy*, vol. 4 (2): 159–219.

Barwise, Jon & John Etchemendy. 1999. *Language, Proof and Logic*. CSLI Publications, New York.

Beall, J. C. 2009. *Spandrels of Truth*. Oxford University Press, Oxford.

Beall, Jc, Michael Glanzberg & David Ripley. 2020. "Liar Paradox." In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta, editor. Metaphysics Research Lab, Stanford University, fall 2020 edn.

Berry, Sharon. 2013. "Default Reasonableness and the Mathoids." *Synthese*, vol. 190 (17): 3695–3713. doi:10.1007/s11229-012-0219-3.

Blake-Turner, Christopher. 2022. "The Hereby-Commit Account of Infer-

ence." *Australasian Journal of Philosophy*, vol. 100 (1): 86–101. doi:10.1080/00048402.2020.1843062.

—. ms./2021. "Against the Constitutivist Construal of Inference."

Blake-Turner, Christopher & Gillian Russell. 2018. "Logical Pluralism Without the Normativity." *Synthese*, 1–19. doi:10.1007/s11229-018-01939-3.

Bledin, Justin. 2014. "Logic Informed." *Mind*, vol. 123 (490): 277–316.

Boer, Steven E. & William G. Lycan. 1980. "Who, Me?" *Philosophical Review*, vol. 89 (3): 427–66.

Boghossian, Paul. 2003. "Blind Reasoning." *Aristotelian Society, Supplementary Volume*, vol. 77 (1): 225–248.

—. 2014. "What is Inference?" *Philosophical Studies*, vol. 169 (1): 1–18.

Bolinger, Dwight. 1968. "Post-posed main phrases: An English rule for the Romance Subjunctive." *Canadian Journal of Linguistics*, vol. 14: 3–30.

Boolos, George. 1993. "Whence the Contradiction?" *Aristotelian Society Supplementary Volume*, vol. 67: 211–233.

Braun, David. 1996. "Demonstratives and Their Linguistic Meanings." *Noûs*, vol. 30 (2): 145–173. doi:10.2307/2216291.

—. 2005. "Empty Names, Fictional Names, Mythical Names." *Noûs*, vol. 39 (4): 596–631. doi:10.1111/j.0029-4624.2005.00541.x.

Broome, John. 1999. "Normative Requirements." *Ratio*, vol. 12 (4): 398–419.

—. 2013. *Rationality Through Reasoning*. Wiley-Blackwell.

—. 2014. "Comments on Boghossian." *Philosophical Studies*, vol. 169 (1): 19–25.

—. 2015. "Synchronic Requirements and Diachronic Permissions." *Canadian Journal of Philosophy*, vol. 45 (5-6): 630–646.

Buckareff, Andrei A. 2005. "How (Not) to Think About Mental Action." *Philosophical Explorations*, vol. 8 (1): 83–89.

Bykvist, Krister & Anandi Hattiangadi. 2007. "Does Thought Imply Ought?" *Analysis*, vol. 67 (296): 277–285.

Byrne, Alex. 2005. "Introspection." *Philosophical Topics*, vol. 33 (1): 79–104. doi:10.5840/philtopics20053312.

—. 2011. "Transparency, Belief, Intention." *Aristotelian Society Supplementary Volume*, vol. 85: 201–21. doi:10.1111/j.1467-8349.2011.00203.x.

—. 2018. *Transparency and Self-Knowledge*. Oxford University Press.

Caie, Michael. 2013. "Rational Probabilistic Incoherence." *Philosophical Review*, vol. 122 (4): 527–575.

Caie, Michael & Dilip Ninan. forthcoming. "First-Person Propositions." *Philosophers' Imprint*.

Cantwell, John. 2008. "Changing the Modal Context." *Theoria*, vol. 74 (4): 331–351. doi:10.1111/j.1755-2567.2008.00028.x.

Cappelen, Herman & Josh Dever. 2013. *The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person*. Oxford University Press.

Carroll, Lewis. 1895. "What the Tortoise Said to Achilles." *Mind*, vol. 4 (14): 278–280.

Cartwright, Richard L. 1994. "Speaking of Everything." *Noûs*, vol. 28 (1): 1–20.

Castañeda, Hector-Neri. 1966. "'He': A Study in the Logic of Self-Consciousness." *Ratio*, vol. 8 (December): 130–57.

—. 1967. "Indicators and Quasi-Indicators." *American Philosophical Quarterly*, vol. 4 (2): 85–100.

—. 1968. "On the Logic of Attributions of Self-Knowledge to Others." *Journal of Philosophy*, vol. 65 (15): 439–456.

Chalmers, David J. 2004. "Epistemic Two-Dimensional Semantics." *Philosophical Studies*, vol. 118 (1-2): 153–226.

Chisholm, R. M. 1963. "Contrary-to-Duty Imperatives and Deontic Logic." *Analysis*, vol. 24 (2): 33–36. doi:10.1093/analys/24.2.33.

Chudnoff, Eli. 2013. *Intuition*. Oxford University Press, Oxford.

—. 2014. "The Rational Roles of Intuition." In *Intuitions*, A. Booth & D. Rowbottom, editors, 9–35. Oxford University Press, Oxford.

Coates, Allen. 2012. "Rational Epistemic Akrasia." *American Philosophical Quarterly*, vol. 49 (2): 113–24.

Cobreros, Pablo. 2011. "Paraconsistent Vagueness: A Positive Argument." *Synthese*, vol. 183 (2): 211–227.

Conant, James. 1991. "The Search for Logically Alien Thought: Descartes, Kant, Frege, and the *Tractatus*." *Philosophical Topics*, vol. 20 (1): 115–180.

Crimmins, Mark. 1995. "Contextuality, Reflexivity, Iteration, Logic." *Philosophical Perspectives*, vol. 9: 381–399. doi:10.2307/2214227.

Cross, Charles & Floris Roelofsen. 2020. "Questions." In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta, editor. Metaphysics Research Lab, Stanford University, fall 2020 edn.

Davies, Martin & Lloyd Humberstone. 1980. "Two Notions of Necessity." *Philosophical Studies*, vol. 38 (1): 1–31.

Descartes, René. 1991. *The Philosophical Writings of Descartes, Vol. 3: Correspondence*. Cambridge: Cambridge University Press.

Deutscher, M. 1969. "A causal account of inferring." In *Contemporary Philosophy in Australia*, R. Brown & C. D. Rollins, editors, 97–118. Allen & Unwin, London.

Devitt, Michael. 2013. "The Myth of the Problematic *De Se*." In *Attitudes de Se: Linguistics, Epistemology, Metaphysics*, A. Capone and N. Feit, editor. CSLI Publications, Stanford.

Dogramaci, Sinan. 2013. "Intuitions for Inferences." *Philosophical Studies*, vol. 165 (2): 371–99.

—. 2015a. "Communist Conventions for Deductive Reasoning." *Noûs*, vol. 49 (4): 776–799. doi:10.1111/nous.12025.

—. 2015b. "Why Is a Valid Inference a Good Inference?" *Philosophy and Phenomenological Research*, vol. 93 (3): 61–96.

—. 2018. "Rational Credence Through Reasoning." *Philosophers' Imprint*, vol. 18.

Douven, Igor. 2013. "The Epistemology of *De Se* Beliefs." In *Attitudes de Se: Linguistics, Epistemology, Metaphysics*, A. Capone and N. Feit, editor. CSLI Publications, Stanford.

Dreier, James. 2009. "Practical Conditionals." In *Reasons for Action*, David Sobel & Steven Wall, editors, 116–133. Cambridge University Press.

Dretske, Fred. 1981. *Knowledge and the Flow of Information*. MIT Press.

Dummett, Michael. 1959. "Truth." *Proceedings of the Aristotelian Society*, vol. 59 (1): 141–62.

—. 1991. *Frege: Philosophy of Mathematics*. Harvard University Press.

—. 1993. "What is Mathematics About?" In *The Seas of Language*, Alexander George, editor, 429–445. Oxford University Press.

Dunn, J. Michael. 1976a. "Intuitive Semantics for First-Degree Entailments and 'Coupled Trees'." *Philosophical Studies*, vol. 29 (3): 149–168. doi:10.1007/BF00373152.

—. 1976b. "A Kripke-Style Semantics for R-Mingle Using a Binary Accessibility Relation." *Studia Logica*, vol. 35 (2): 163–172. doi:10.1007/BF02120878.

Dutilh Novaes, Catarina. 2020. *The Dialogical Roots of Deduction: Historical, Cognitive, and Philosophical Perspectives on Reasoning*. Cambridge University Press.

Edgington, Dorothy. 1996. "Lowe on Conditional Probability." *Mind*, vol. 105 (420): 617–630. doi:10.1093/mind/105.420.617.

—. 2020. "Indicative Conditionals." In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta, editor. Metaphysics Research Lab, Stanford University, fall 2020 edn.

Efird, David & Tom Stoneham. 2008. "What is the Principle of Recombination?" *Dialectica*, vol. 62 (4): 483–494.

Etchemendy, John. 1990. *The Concept of Logical Consequence*. Harvard University Press.

—. 2008. "Reflections on Consequence." In *New Essays on Tarski and Philosophy*, Douglas Patterson, editor, 263–299. Oxford University Press.

Evans, Gareth. 1979. "Reference and Contingency." *The Monist*, vol. 62 (2): 161–189.

—. 1982. *The Varieties of Reference*. 137. Oxford University Press.

Farkas, Donka. 1985. *Intensional Descriptions and the Romance Subjunctive*. Garland, New York.

Feferman, Solomon. 1984. "Toward Useful Type-Free Theories. I." *Journal of Symbolic Logic*, vol. 49 (1): 75–111.

Feit, Neil. 2010. "Selfless Desires and the Property Theory of Content." *Australasian Journal of Philosophy*, vol. 88 (3): 489–503. doi:10.1080/00048400903193361.

Ferrari, Filippo, Ben Martin & Maria Paola Fogliani Sforza. 2023. "Anti-Exceptionalism About Logic: An Overview." *Synthese*, vol. 201 (2): 1–9. doi:10.1007/s11229-023-04082-w.

Field, Hartry. 1989. *Realism, Mathematics and Modality*. Blackwell, Oxford.

—. 1994. "Deflationist Views of Meaning and Content." *Mind*, vol. 103 (411): 249–285.

—. 2009. "What is the Normative Role of Logic?" *Aristotelian Society Supplementary Volume*, vol. 83 (1): 251–268.

—. 2015. "What Is Logical Validity?" In *Foundations of Logical Consequence*, Colin R. Caret & Ole T. Hjortland, editors, 33–70. Oxford University Press.

Field, Hartry H. 2008. *Saving Truth From Paradox*. Oxford University Press.

Fine, Kit. 1975. "Vagueness, Truth and Logic." *Synthese*, vol. 30 (3-4): 265–300.

Finlay, Stephen. 2004. "The Conversational Practicality of Value Judgment." *The Journal of Ethics*, vol. 8: 205–223.

—. 2014. *A Confusion of Tongues*. Oxford University Press, New York.

Fintel, von & Sabine Iatridou. 2023. "Prolegomena to a Theory of X-Marking." *Linguistics and Philosophy*, vol. 46 (6): 1467–1510. doi:10.1007/s10988-023-09390-5.

von Fintel, Kai. 1994. *Restrictions on Quantifier Domains*. Ph.D. thesis, University of Massachusetts at Amherst.

—. 1999. "NPI Licensing, Strawson Entailment, and Context Dependency." *Journal of Semantics*, vol. 16 (2): 97–148. doi:10.1093/jos/16.2.97.

—. 2004. "Would You Believe It? The King of France is Back! (Presuppositions and Truth-Value Intuitions)." In *Descriptions and Beyond*, Marga Reimer & Anne Bezuidenhout, editors. Clarendon Press.

von Fintel, Kai & Anthony S. Gillies. 2010. "Must . . . Stay . . . Strong!" *Natural Language Semantics*, vol. 18 (4): 351–383. doi:10.1007/s11050-010-9058-2.

Fodor, Jerry A. 1975. *The Language of Thought*. Harvard University Press.

Forrest, Peter & D. M. Armstrong. 1984. "An Argument Against David Lewis' Theory of Possible Worlds." *Australasian Journal of Philosophy*, vol. 62 (2): 164–168.

Forrester, James William. 1984. "Gentle Murder, or the Adverbial Samaritan." *Journal of Philosophy*, vol. 81 (4): 193–197. doi:10.2307/2026120.

van Frassen, Bas C. 1968. "Presupposition, implication, and self-reference." *Journal of Philosophy*, vol. 65 (5): 136–152.

—. 1970. "Truth and paradoxical consequence." In *The Paradox of the Liar*, Robert L. Martin, editor, 13–23. Atascadero, Ridgeview.

Frege, Gottlob. 1879?/1983. "Logik." In *Nachgelassene Schriften und wissenschaftlicher Briefwechsel*, vol. 1, 1–8. Felix Meiner Verlag, Hamburg.

—. 1906. "Über die Grundlagen der Geometrie II." *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 15: 377–403.

—. 1918/1997. "Thought." In *The Frege Reader*, Michael Beaney, editor. Blackwell.

Friedman, Jane. 2013a. "Question-Directed Attitudes." *Philosophical Perspectives*, vol. 27 (1): 145–174.

—. 2013b. "Suspended Judgment." *Philosophical Studies*, vol. 162 (2): 165–181.

—. 2017a. "Inquiry and Belief." *Noûs*, vol. 53 (2): 296–315. doi:10.1111/nous.12222.

—. 2017b. "Why Suspend Judging?" *Noûs*, vol. 51 (2): 302–326.

Gazdar, G. 1979. *Pragmatics: Implicature, Presupposition, and Logical Form*. Academic Press, New York.

Gibbons, John. 2009. "Reason in Action." In *Mental Actions*, Lucy O'Brien & Matthew Soteriou, editors, 72–94. Oxford University Press.

Glanzberg, Michael. 2003. "Against Truth-Value Gaps." In *Liars and Heaps*, J. C. Beall, editor, 151–94. Oxford University Press.

—. 2015. "Logical Consequence and Natural Language." In *Foundations of Logical Consequence*, Colin Caret & Ole Hjortland, editors, 71–120. Oxford University Press.

Glanzberg, Michael & Susanna Siegel. 2006. "Presupposition and Policing in Complex Demonstratives." *Noûs*, vol. 40 (1): 1–42.

Goldstein, Simon & Paolo Santorio. 2021. "Probability for Epistemic Modalities." *Philosophers' Imprint*, vol. 21 (33).

Gómez-Torrente, Mario. 2008. "Are There Model-Theoretic Logical Truths That Are Not Logically True?" In *New Essays on Tarski and Philosophy*, Douglas Patterson, editor, 340–368. Oxford University Press.

Greco, Daniel. 2014. "A Puzzle About Epistemic Akrasia." *Philosophical Studies*, vol. 167 (2): 201–219. doi:10.1007/s11098-012-0085-3.

Groenendijk, Jerome & Martin Stokhof. 1984. *Studies int he Semantics of Questions and the Pragmatics of Answers*. Ph.D. thesis, University of Amsterdam.

—. 1997. "Questions." In *Handbook of Logic and Language*, J. van Bentham & A. ter Meulen, editors, 1055–1124. Elsevier Science, Amsterdam.

Gupta, Anil & Nuel Belnap. 1993. *The Revision Theory of Truth*. MIT Press, Cambridge.

Hamblin, Charles L. 1973. "Questions in Montague English." *Foundations of Language*, vol. 10 (1): 41–53.

Hanna, Robert. 2006. "Rationality and the Ethics of Logic." *Journal of Philosophy*, vol. 103 (2): 67–100. doi:10.5840/jphil2006103235.

Hanson, William H. 1997. "The Concept of Logical Consequence." *Philosophical Review*, vol. 106 (3): 365–409.

—. 2006. "Actuality, Necessity, and Logical Truth." *Philosophical Studies*, vol. 130 (3): 437–459.

Harman, Gilbert. 1984. "Logic and Reasoning." *Synthese*, vol. 60 (1): 107–127.

—. 1986. *Change in View*. MIT Press.

Hausser, R. & D. Zaeffer. 1979. "Questions and Answers in a Context-Dependent Montague grammar." In *Formal Semantics and Pragmatics for Natural Languages*, F. Guenthner & S. Schmidt, editors, 339–358. Reidel, Dordrecht.

Hawthorne, John, Daniel Rothschild & Levi Spectre. 2016. "Belief is Weak." *Philosophical Studies*, vol. 173 (5): 1393–1404. doi:10.1007/s11098-015-0553-7.

Hazlett, Allan. 2013. *A Luxury of the Understanding: On the Value of True Belief*. Oxford University Press, Oxford.

Hedden, Brian. 2015a. *Reasons Without Persons: Rationality, Identity, and Time*. Oxford University Press UK.

—. 2015b. "Time-Slice Rationality." *Mind*, vol. 124 (494): 449–491. doi:10.1093/mind/fzu181.

Heim, Irene. 1992. "Presupposition Projection and the Semantics of Attitude Verbs." *Journal of Semantics*, vol. 9 (3): 183–221.

Heim, Irene & Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Blackwell.

Hieronymi, Pamela. 2009. "Two Kinds of Agency." In *Mental Action*, Lucy O'Brien & Matthew Soteriou, editors, 138–162. Oxford University Press.

Hlobil, Ulf. 2014. "Against Boghossian, Wright and Broome on Inference." *Philosophical Studies*, vol. 167 (2): 419–429.

—. 2015. "There Are Diachronic Norms of Rationality." *Thought: A Journal of Philosophy*, vol. 4 (1): 38–45. doi:10.1002/tht3.155.

—. 2016a. "Chains of Inferences and the New Paradigm in the Psychology of Reasoning." *Review of Philosophy and Psychology*, vol. 7 (1): 1–16. doi:10.1007/s13164-015-0230-y.

—. 2016b. *What is Inference? Or the Force of Reasoning*. Ph.D. thesis, University of Pittsburgh.

—. 2019. "Inferring by Attaching Force." *Australasian Journal of Philosophy*, vol. 97 (4): 701–714. doi:10.1080/00048402.2018.1564060.

Holton, Richard. 2014. "Intention as a Model for Belief." In *Rational and Social Agency: Essays on the Philosophy of Michael Bratman*, Manuel Vargas & Gideon Yaffe, editors. Oxford University Press.

Hurley, Susan L. 1989. *Natural Reasons: Personality and Polity*. Oxford University Press.

Husserl, Edmund. 1969. *Formal and Transcendental Logic*. The Hague: Martinus Nijhoff.

Hyde, Dominic & Mark Colyvan. 2008. "Paraconsistent Vagueness: Why Not?" *Australasian Journal of Logic*, vol. 6 (3): 107–121.

Kamp, Hans. 1971. "Formal Properties of 'Now'." *Theoria*, vol. 37 (3): 227–273. doi:10.1111/j.1755-2567.1971.tb00071.x.

Kant, Immanuel. 1992. *Lectures on Logic*. Cambridge University Press.

Kaplan, David. 1989a. "Afterthoughts." In *Themes From Kaplan*, J. Almog, J. Perry & H. Wettstein, editors, 565–614. Oxford University Press.

—. 1989b. "Demonstratives." In *Themes From Kaplan*, Joseph Almog, John Perry & Howard Wettstein, editors, 481–563. Oxford University Press.

Karttunen, Lauri. 1973. "Presuppositions of compound sentences." *Linguistic Inquiry*, vol. 4 (2): 167–193.

—. 1977. "Syntax and Semantics of Questions." *Linguistics and Philosophy*, vol. 1 (1): 3–44. doi:10.1007/BF00351935.

Keenan, D. & D. Westerstahl. 2011. "Generalized Quantifiers in Linguistics and Logic." In *Handbook of Logic and Language*, Johan Van Benthem & Alice Ter Meulen, editors, 859–910. Elsevier.

Kennedy, Christopher. 2007. "Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives." *Linguistics and Philosophy*, vol. 30 (1): 1–45. doi:10.1007/s10988-006-9008-0.

Kimhi, Irad. 2018. *Thinking and Being*. Harvard University Press.

Kind, Amy. 2001. "Putting the Image Back in Imagination." *Philosophy and Phenomenological Research*, vol. 62 (1): 85–110.

Kneal, William & Martha Kneale. 1962. *The Development of Logic*. Clarendon Press, Oxford.

Kolodny, Niko & John MacFarlane. 2010. "Ifs and Oughts." *Journal of Philosophy*, vol. 107 (3): 115–143. doi:10.5840/jphil20101073l0.

Korcz, Keith Allen. 2021. "The Epistemic Basing Relation." In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta, editor. Metaphysics Research Lab, Stanford University, Spring 2021 edn.

Kratzer, Angelika. 1977. "What 'Must' and 'Can' Must and Can Mean." *Linguistics and Philosophy*, vol. 1 (3): 337–355. doi:10.1007/BF00353453.

—. 1981. "The Notional Category of Modality." In *Words, Worlds, and Contexts: New Approaches in Word Semantics*, Hans-Jürgen Eikmeyer & Hannes Rieser, editors, 39–74. W. De Gruyter, Berlin.

—. 1986. "Conditionals." *Chicago Linguistics Society*, vol. 22 (2): 1–15.

Kreisel, Georg. 1967. "Informal Rigour and Completeness Proofs." In *Problems in the Philosophy of Mathematics*, Imre Lakatos, editor, 138–157. North-Holland.

Kripke, Saul. 2013. *Reference and Existence: The John Locke Lectures*. Oxford University Press.

Kripke, Saul A. 1963. "Semantical Considerations on Modal Logic." *Acta Philosophica Fennica*, vol. 16 (1963): 83–94.

—. 1975. "Outline of a Theory of Truth." *Journal of Philosophy*, vol. 72 (19): 690–716.

—. 1980. *Naming and Necessity*. Harvard University Press.

—. forthcoming. "The Question of Logic." *Mind*. doi:10.1093/mind/fzad008.

Kung, Peter. 2010. "Imagining as a Guide to Possibility." *Philosophy and Phenomenological Research*, vol. 81 (3): 620–663. doi:10.1111/j.1933-1592.2010.00377.x.

Lance, Mark Norris. 1995. "Subjective Probability and Acceptance." *Philosophical Studies*, vol. 77 (1): 147–179. doi:10.1007/BF00996316.

Lappin, Shalom. 1981. *Sorts, Ontology, and Metaphor: The Semantics of Sortal Structure*. W. De Gruyter.

Leech, Jessica. 2015. "Logic and the Laws of Thought." *Philosophers' Imprint*, vol. 15.

Lewis, David. 1979. "Attitudes de Dicto and de Se." *Philosophical Review*, vol. 88 (4): 513–543.

Lewis, David. 1982. "Logic for Equivocators." *Noûs*, vol. 16 (3): 431–441.

—. 1986. *On the Plurality of Worlds*. Blackwell.

Lewis, David K. 1975. "Adverbs of Quantification." In *Formal Semantics of Natural Language*, Edward L. Keenan, editor, 178–188. Cambridge University Press.

—. 1976. "The Paradoxes of Time Travel." *American Philosophical Quarterly*, vol. 13 (2): 145–152.

—. 1980. "Index, Context, and Content." In *Philosophy and Grammar*, S. Kanger & S. Öhman, editors, 79–106. D. Reidel, Holland.

Locke, John. 1690/1979. *An Essay Concerning Human Understanding*. Oxford University Press.

Longuenesse, Béatrice. 2005. *Kant on the Human Standpoint*. Cambridge University Press.

Lu-Adler, Huaping. 2017. "Kant and the Normativity of Logic." *European Journal of Philosophy*, vol. 25 (2): 207–230. doi:10.1111/ejop.12242.

Łukasiewicz, Jan. 1910/1979. "Aristotle on the Law of Contradiction." In *Articles on Aristotle Vol.3: Metaphysics*, Jonathan Barnes, Malcolm Schofield & Richard Sorabji, editors, 50–62. Duckworth, London.

—. 1930/1970. "Philosophical Remarks on Many-Valued Systems of Propositional Logic." In *Jan Łukasiewicz: Selected Works*. Amsterdam: North Holland.

MacFarlane, John. 2002. "Frege, Kant, and the Logic in Logicism." *Philosophical Review*, vol. 111 (1): 25–65. doi:10.1215/00318108-111-1-25.

—. 2014. *Assessment Sensitivity: Relative Truth and its Applications*. Oxford University Press.

MacFarlane, John. ms/2004. "In what sense (if any) is logic normative for thought?" Unpublished. Delivered at the American Philosophical Association Central Division meeting.

Magidor, Ofra. 2015. "The Myth of the De Se." *Philosophical Perspectives*, vol. 29: 249—283.

Maier, Emar. 2016. "Why My I is Your You: On the Communication of de Se Attitudes." In *About Oneself: De Se Thought and Communication*, Manuel Garcia-Carpintero & Stephan Torre, editors. Oxford University Press.

Mandelkern, Matthew. 2020. "A Counterexample to Modus Ponenses." *Journal of Philosophy*, vol. 117 (6): 315–331. doi:10.5840/jphil2020117619.

Marcus, Eric. 2012. *Rational Causation*. Harvard University Press.

—. 2020. "Inference as Consciousness of Necessity." *Analytic Philosophy*, vol. 61 (4): 304–322. doi:10.1111/phib.12153.

—. 2021. *Belief, Inference, and the Self-Conscious Mind*. Oxford University Press.

Markie, Peter J. 1984. "De Dicto and de Se." *Philosophical Studies*, vol. 45 (2): 231–237. doi:10.1007/BF00372482.

Marushak, Adam & James R. Shaw. ms./2020. "Epistemics and Emotives."

May, Robert. 1985. *Logical Form: Its Structure and Derivation*. MIT Press.

McGee, Vann. 1985. "A Counterexample to Modus Ponens." *Journal of Philosophy*, vol. 82 (9): 462–471. doi:jphil198582937.

—. 1992. "Two Problems with Tarski's Theory of Consequence." *Proceedings of the Aristotelian Society*, vol. 92: 273–292.

McHugh, Conor & Jonathan Way. 2016. "Against the Taking Condition." *Philosophical Issues*, vol. 26 (1): 314–331. doi:10.1111/phis.12074.

—. 2018. "What is Reasoning?" *Mind*, vol. 127 (505): 167–196. doi:10.1093/mind/fzw068.

Mele, Alfred. 2009. "Mental Action: A Case Study." In *Mental Actions*, Lucy O'Brien & Matthew Soteriou, editors, 17–39. Oxford University Press.

Merritt, Melissa McBay. 2015. "Varieties of Reflection in Kant's Logic." *British Journal for the History of Philosophy*, vol. 23 (3): 478–501. doi:10.1080/09608788.2015.1018129.

Millikan, Ruth G. 1984. *Language, Thought and Other Biological Categories*. MIT Press.

Millikan, Ruth Garrett. 1990. "The Myth of the Essential Indexical." *Noûs*, vol. 24 (5): 723–734. doi:10.2307/2215811.

Moss, Sarah. 2018. *Probabilistic Knowledge*. Oxford University Press.

—. 2019. "Full Belief and Loose Speech." *Philosophical Review*, vol. 128 (3): 255–291. doi:10.1215/00318108-7537270.

Neta, Ram. 2013. "What is an Inference." *Philosophical Issues*, vol. 23 (1): 388–407. doi:10.1111/phis.12020.

Ninan, Dilip. 2010a. "De Se Attitudes: Ascription and Communication." *Philosophy Compass*, vol. 5 (7): 551–567. doi:10.1111/j.1747-9991.2010.00290.x.

—. 2010b. "Semantics and the Objects of Assertion." *Linguistics and Philosophy*, vol. 33 (5): 355–380.

Ninan, Dilip. 2016. "What is the Problem of De Se Attitudes?" In *About Oneself: De Se Thought and Communication*, Stephan Torre & Manuel Garcia-Carpintero, editors, 86–120. Oxford University Press.

Nir, Gilad. 2021. "Are Rules of Inference Superfluous? Wittgenstein Vs. Frege and Russell." *Teorema: International Journal of Philosophy*, vol. 40 (2): 45–61.

Nolan, Daniel. 1996. "Recombination Unbound." *Philosophical Studies*, vol. 84 (2-3): 239–262.

—. 2006. "Selfless Desires." *Philosophy and Phenomenological Research*, vol. 73 (3): 665–679. doi:10.1111/j.1933-1592.2006.tb00553.x.

Noordhof, Paul. 2002. "Imagining Objects and Imagining Experiences." *Mind and Language*, vol. 17 (4): 426–455.

Norton, John D. 2003. "A Material Theory of Induction." *Philosophy of Science*, vol. 70 (4): 647–670. doi:10.1086/378858.

—. 2014. "A Material Dissolution of the Problem of Induction." *Synthese*, vol. 191 (4): 1–20. doi:10.1007/s11229-013-0356-3.

—. 2021. *The Material Theory of Induction*. BSPS Open.

Nunez, Tyke. 2018. "Logical Mistakes, Logical Aliens, and the Laws of Kant's Pure General Logic." *Mind*, 1149–1180. doi:10.1093/mind/fzy027.

Over, D. E. 1987. "Assumptions and the Supposed Counterexamples to Modus Ponens." *Analysis*, vol. 47 (3): 142. doi:10.1093/analys/47.3.142.

Owens, David. 2002. "Epistemic Akrasia." *The Monist*, vol. 85 (3): 381–397. doi:10.5840/monist200285316.

Oza, Manish. 2020. "The Value of Thinking and the Normativity of Logic." *Philosophers' Imprint*, vol. 20 (25): 1–23.

Padró, Romina. 2015. *What the Tortoise Said to Kripke: the Adoption Problem and the Epistemology of Logic*. Ph.D. thesis, City University of New York.

Pap, Arthur. 1958. *Semantics and Necessary Truth*. Yale University Press, New Haven.

Parfit, Derek. 1981. "What We Together Do."

Peacocke, Christopher. 1985. "Imagination, Experience, and Possibility." In *Essays on Berkeley: A Tercentennial Celebration*, John Foster & Howard Robinson, editors. Oxford University Press.

—. 2008. "Mental Action and Self-Awareness." In *Mental Action*, Lucy F. O'Brien & Matthew Soteriou, editors, 358–376. Oxford University Press.

Peirce, Charles Sanders. 1905. "Issues of Pragmaticism." *The Monist*, vol. 15 (4): 481–99.

Perry, John. 1979. "The Problem of the Essential Indexical." *Noûs*, vol. 13 (December): 3–21.

Pettit, Philip & Michael Smith. 1996. "Freedom in Belief and Desire." *Journal of Philosophy*, vol. 93 (9): 429–449. doi:jphil199693915.

Plantinga, Alvin. 1993. *Warrant and Proper Function*. Oxford University Press.

Podgorski, Abelard. 2016. "A Reply to the Synchronist." *Mind*, vol. 125 (499): 859–871. doi:10.1093/mind/fzv153.

—. 2017. "Rational Delay." *Philosophers' Imprint*, vol. 17.

Priest, Graham. 1979a. "The Logic of Paradox." *Journal of Philosophical Logic*, vol. 8 (1): 219–241. doi:10.1007/BF00258428.

—. 1979b. "Two Dogmas of Quineanism." *Philosophical Quarterly*, vol. 29 (117): 289–301.

—. 1984. "Logic of Paradox Revisited." *Journal of Philosophical Logic*, vol. 13 (2): 153–179.

—. 2006. *In Contradiction: A Study of the Transconsistent*. Oxford University Press.

—. 2016. "Thinking the Impossible." *Philosophical Studies*, vol. 173 (10): 2649–2662. doi:10.1007/s11098-016-0668-5.

Putnam, Hilary. 1994. "Rethinking Mathematical Necessity." In *Words and Life*, J. Conant, editor, 245–263. Harvard University Press, Cambridge, MA.

Quine, Willard V. O. 1970/86. *Philosophy of Logic*. Harvard University Press, Cambridge, MA, second edn.

Rabern, Brian. 2012. "Against the Identification of Assertoric Content with Semantic Value." *Synthese*, vol. 189 (1): 75–96.

—. 2013. "Monsters in Kaplan's Logic of Demonstratives." *Philosophical Studies*, vol. 164 (2): 393–404.

Rayo, Agustín & Gabriel Uzquiano. 2006. *Absolute Generality*. Oxford University Press.

Raz, Joseph. 2009. "Reasons : Practical and Adaptive." In *Reasons for Action*, David Sobel & Steven Wall, editors. Cambridge University Press.

Regan, Donald H. 1980. *Utilitarianism and Co-Operation*. Oxford University Press.

Rescorla, Michael. 2019. "The Language of Thought Hypothesis." In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta, editor. Metaphysics Research Lab, Stanford University, summer 2019 edn.

Restall, Greg. 2006. *Logic: An Introduction*. Routledge, New York.

Rothschild, Daniel. 2012. "Expressing Credences." *Proceedings of the Aristotelian Society*, vol. 112 (1pt1): 99–114. doi:10.1111/j.1467-9264.2012.00327.x.

Rumfitt, Ian. 2015. *The Boundary Stones of Thought: An Essay in the Philosophy of Logic*. Oxford University Press.

Russell, Bertrand. 1920/1988. "The Nature of Inference." In *The Collected Papers of Bertrand Russell, vol.9: Essays on Language, Mind, and Matter, 1919–26*, J. G. Slater & B. Frohmann, editors. Unwin Hyman, London.

Russell, G. 2012. "Lessons From the Logic of Demonstratives: What Indexicality Teaches Us About Logic and Vice Versa." In *New Waves in Philosophical Logic*, Greg Restall & Gillian Kay Russell, editors. Palgrave-Macmillan.

Russell, Gillian. 2017. "Logic Isn't Normative." *Inquiry: An Interdisciplinary Journal of Philosophy*, 1–18. doi:10.1080/0020174x.2017.1372305.

Russell, Gillian Kay. 2008. *Truth in Virtue of Meaning*. Oxford University Press.

Russell, Jeffrey Sanford & John Hawthorne. 2018. "Possible Patterns." *Oxford Studies in Metaphysics*, vol. 11.

Sagi, Gil. 2014. "Models and Logical Consequence." *Journal of Philosophical Logic*, vol. 43 (5): 943–964.

Sainsbury, Roy M. 1991/2001. *Logical Forms: An Introduction to Philosophical Logic*. Blackwell, Oxford, second edn.

—. 2002. "What Logic Should We Think With?" In *Logic, Thought and Language*, Anthony O'Hear, editor, 1–17. Cambridge University Press, Cambridge.

Salmon, Nathan. 1993. "Relative and Absolute Apriority." *Philosophical Studies*, vol. 69 (1): 83–100.

—. 2002. "Demonstrating and Necessity." *Philosophical Review*, vol. III (4): 497–537. doi:10.1215/00318108-111-4-497.

Salmon, Wesley. 1963/73. *Logic*. Prentice Hall, Englewood Cliffs, NJ, second edn.

Sánchez-Miguel, Manuel García-Carpintero. 1992. "The Grounds for the Model-Theoretic Account of the Logical Properties." *Notre Dame Journal of Formal Logic*, vol. 34 (1): 107–131.

van der Sandt, R.A. 1988. *Context and Presupposition*. Routledge, London.

Santorio, Paolo. 2012. "Reference and Monstrosity." *Philosophical Review*, vol. 121 (3): 359–406.

—. 2022. "Trivializing Informational Consequence." *Philosophy and Phenomenological Research*, vol. 104 (2): 297–320. doi:10.1111/phpr.12745.

Scanlon, Thomas. 1998. *What We Owe to Each Other*. Harvard University Press.

Schechter, Joshua. 2019. "Small Steps and Great Leaps in Thought: The Epistemology of Basic Deductive Rules." In *Reasoning: New Essays on Theoretical and Practical Thinking*, Magdalena Balcerak Jackson & Brendan Balcerak Jackson, editors. Oxford: Oxford University Press.

Schiffer, Stephen. 2003. *The Things We Mean*. Oxford University Press.

Schlenker, P. 2003. "A plea for monsters." *Linguistics and Philosophy*, vol. 26 (1): 29–120.

Schroeder, Mark. 2015. *Expressing Our Attitudes: Explanation and Expression in Ethics, Vol.2*. Oxford University Press UK.

Schulz, Moritz. 2010. "Epistemic Modals and Informational Consequence." *Synthese*, vol. 174 (3): 385–395. doi:10.1007/s11229-009-9461-8.

Sennet, Adam. 2016. "Ambiguity." In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta, editor. Metaphysics Research Lab, Stanford University, spring 2016 edn.

Setiya, Kieran. 2013. "Epistemic Agency: Some Doubts." *Philosophical Issues*, vol. 23 (1): 179–198.

Shapiro, Stewart. 1998. "Logical Consequence: Models and Modality." In *The Philosophy of Mathematics Today*, Matthias Schirn, editor, 131–156. Clarendon Press.

Shaw, James R. 2014. "What is a Truth-Value Gap?" *Linguistics and Philosophy*, vol. 37 (6): 503–534.

—. 2015. "Anomaly and Quantification." *Noûs*, vol. 49 (1): 147–176.

—. 2016. "Magidor on Anomaly and Truth-Value Gaps." *Inquiry: An Interdisci-*

plinary Journal of Philosophy, vol. 59 (5): 513–528. doi:10.1080/0020174X.2016.1184841.

—. 2020. "De Se Exceptionalism and Frege Puzzles." *Ergo*, vol. 6: 1057–1086. doi:10.3998/ergo.12405314.0006.037.

—. 2021/ms. "Paradoxical Mentality and Attitudinal Embedding."

—. 2023. *Wittgenstein on Rules: Justification, Grammar, and Agreement*. Oxford University Press.

Sider, Theodore. 2005. "Another Look at Armstrong's Combinatorialism." *Noûs*, vol. 39 (4): 679–695.

Siegel, Susanna. 2017. *The Rationality of Perception*. Oxford University Press.

—. 2019. "Inference Without Reckoning." In *Reasoning: New Essays on Theoretical and Practical Thinking*, Brendan Balcerak Jackson & Magdalena Balcerak Jackson, editors, 15–31. Oxford University Press.

Sinnott-Armstrong, Walter, James Moor & Robert Fogelin. 1986. "A Defense of Modus Ponens." *Journal of Philosophy*, vol. 83 (5): 296. doi:10.2307/2026144.

Soames, Scott. 1982. "How Presuppositions Are Inherited: A Solution to the Projection Problem." *Linguistic Inquiry*, vol. 13: 483–545.

—. 1989. "Presupposition." In *Handbook of Philosophical Logic*, D. Gabbay & F. Guenthner, editors. Kluwer, Dordrecht.

—. 1999. *Understanding Truth*. Oxford University Press USA.

—. 2005. *Reference and Description: The Case Against Two-Dimensionalism*. Princeton: Princeton University Press.

Spencer, Cara. 2007. "Is There a Problem of the Essential Indexical?" In *Situating Semantics: Essays on the Philosophy of John Perry*, M. O'Rourke & C. Washington, editors. MIT Press, Cambridge.

de Spinoza, Benedictus. 1985. *The Collected Works of Spinoza: Vol I*. Princeton University Press.

Staffel, Julia. 2013. "Can There Be Reasoning with Degrees of Belief?" *Synthese*, vol. 190 (16): 3535–3551. doi:10.1007/s11229-012-0209-5.

Stalnaker, Robert. 1978. "Assertion." *Syntax and Semantics (New York Academic Press)*, vol. 9: 315–332.

—. 1984. *Inquiry*. Bradford Books, MIT Press.

Stalnaker, Robert C. 1968. "A Theory of Conditionals." In *Studies in Logical Theory (American Philosophical Quarterly Monographs 2)*, Nicholas Rescher, editor, 98–112. Oxford: Blackwell.

—. 1981. "Indexical Belief." *Synthese*, vol. 49 (1): 129–151.

Stang, Nicholas. 2014. "Kant, Bolzano, and the Formality of Logic." In *The New Anti-Kant*, Sandra Lapointe & Clinton Tolley, editors, 193–234. Palgrave Macmillan.

Stanley, Jason. 1997. "Rigidity and Content." In *Language, Thought, and Logic: Essays in honor of Michael Dummett*, 131–156. Clarendon Press, Oxford.

Stanley, Jason & Zoltán Gendler Szabó. 2000. "On Quantifier Domain Restriction." *Mind and Language*, vol. 15 (2&3): 219–61.

Steinberger, Florian. 2016. "Explosion and the Normativity of Logic." *Mind*, vol. 125 (498): 385–419.

—. 2019a. "Consequence and Normative Guidance." *Philosophy and Phenomenological Research*, vol. 98 (2): 306–328. doi:10.1111/phpr.12434.

—. 2019b. "Three Ways in Which Logic Might Be Normative." *Journal of Philosophy*, vol. 116 (1): 5–31. doi:10.5840/jphil2019116111.

Strawson, Galen. 2003. "Mental Ballistics or the Involuntariness of Spontaniety." *Proceedings of the Aristotelian Society*, vol. 103 (3): 227–257.

Strawson, P. F. 1952. *Introduction to Logical Theory*. Routledge.

Stroud, Barry. 1979. "Inference, Belief, and Understanding." *Mind*, vol. 88 (350): 179–196.

Thomason, R. H. 1972. "A Semantic Theory of Sortal Incorrectness." *Journal of Philosophical Logic*, vol. 1 (2): 209–258.

Thomson, Judith Jarvis. 1965. "Reasons and Reasoning." In *Philosophy in America*, Max Black, editor, 282–303. Cornell University Press, Ithaca, N.Y.

—. 2008. *Normativity*. Open Court Press, Chicago.

Titelbaum, Michael G. 2014. *Quitting Certainties: A Bayesian Framework Modeling Degrees of Belief*. Oxford University Press.

Tolley, Clinton. 2006. "Kant on the Nature of Logical Laws." *Philosophical Topics*, vol. 34 (1/2): 371–407. doi:10.5840/philtopics2006341/214.

Torre, Stephan. 2018. "In Defense of De Se Content." *Philosophy and Phenomenological Research*, vol. 97 (1): 172–189.

Travis, Charles. 2019. "Where Words Fail." In *The Logical Alien: Conant and his Critics*, Sofia Miguens, editor, 222–281. Harvard University Press, Cambridge.

Turner, Jason. 2010. "Fitting Attitudes de Dicto and de Se." *Noûs*, vol. 44 (1): 1–9. doi:10.1111/j.1468-0068.2009.00728.x.

Valaris, Markos. 2011. "Transparency as Inference: Reply to Alex Byrne." *Proceedings of the Aristotelian Society*, vol. 111 (2pt2): 319–324. doi:10.1111/j.1467-9264.2011.00312.x.

—. 2014. "Reasoning and Regress." *Mind*, vol. 123 (489): 101–127.

—. 2017. "What The Tortoise Has To Say About Diachronic Rationality." *Pacific Philosophical Quarterly*, vol. 98 (S1): 293–307.

—. 2020. "Reasoning, Defeasibility, and the Taking Condition." *Philosophers' Imprint*, vol. 20 (28): 1–16.

Veltman, Frank. 1996. "Defaults in Update Semantics." *Journal of Philosophical Logic*, vol. 25 (3): 221–261. doi:10.1007/BF00248150.

Visser, Albert. 2004. "Semantics and the Liar Paradox." In *Handbook of Philosophical Logic*, D. Gabbay & F. Guethner, editors, vol. 11, 149–240. Springer, second edn.

Wang, Jennifer. 2013. "From Combinatorialism to Primitivism." *Australasian Journal of Philosophy*, vol. 91 (3): 535–554.

Weatherson, Brian. 2008. "Deontology and Descartes's Demon." *Journal of Philosophy*, vol. 105 (9): 540–569. doi:10.5840/jphil2008105932.

Weber, Z. 2010. "A Paraconsistent Model of Vagueness." *Mind*, vol. 119 (476): 1025–1045.

Weisberg, Jonathan. 2013. "Knowledge in Action." *Philosophers' Imprint*, vol. 13.

Willer, Malte. 2012. "A Remark on Iffy Oughts." *Journal of Philosophy*, vol. 109 (7): 449–461. doi:10.5840/jphil2012109719.

Williamson, Timothy. 1994. *Vagueness*. Routledge.

—. 2000. *Knowledge and its Limits*. Oxford University Press.

—. 2003. "Everything." *Philosophical Perspectives*, vol. 17 (1): 415–465.

—. 2013. *Modal Logic as Metaphysics*. Oxford University Press.

Winters, Barbara. 1983. "Inferring." *Philosophical Studies*, vol. 105 (4): 201–220.

Wittgenstein, Ludwig. 1922. *Tractatus Logico-Philosophicus*. Dover Publications.

Wright, Crispin. 2014. "Comment on Paul Boghossian, "What is Inference"." *Philosophical Studies*, vol. 169 (1): 27–37.

Yalcin, Seth. 2007. "Epistemic Modals." *Mind*, vol. 116 (464): 983–1026. doi:10.1093/mind/fzm983.

—. 2010. "Probability Operators." *Philosophy Compass*, vol. 5 (11): 916–37. doi:10.1111/j.1747-9991.2010.00360.x.

—. 2011. "Nonfactualism About Epistemic Modality." In *Epistemic Modality*, Andy Egan & B. Weatherson, editors. Oxford University Press.

—. 2012a. "Context Probabilism." In *Logic, Language, and Meaning: 18th Amsterdam Colloquium, Amsterdam, The Netherlands, December 19-21, 2011, Revised Selected Papers*, M. Aloni, Maria Aloni, Vadim Kimmelmann, Floris Roelofson, Galit W. Sassoon, Katrin Schulz & Matthijs Westera, editors, 12–21. Berlin: Springer.

—. 2012b. "A Counterexample to Modus Tollens." *Journal of Philosophical Logic*, vol. 41 (6): 1001–1024. doi:10.1007/s10992-012-9228-4.

—. 2014. "Semantics and Metasemantics in the Context of Generative Grammar." In *Metasemantics: New Essays on the Foundations of Meaning*, Alexis Burgess & Brett Sherman, editors, 17–54. Oxford University Press, Oxford.

—. 2015. "Actually, Actually." *Analysis*, vol. 75 (2): 185–191.

Zalta, Edward N. 1988. "Logical and Analytic Truths That Are Not Necessary." *Journal of Philosophy*, vol. 85 (2): 57–74.

Zeevat, Henk. 1992. "Presupposition and Accommodation in Update Semantics." *Journal of Semantics*, vol. 9 (4): 379–412.

Ziff, Paul. 1960. *Semantic Analysis*. Cornell University Press, Ithaca.