Time Travel, Anaphora, and Attitudes

James R. Shaw

Time travelers supply their share of metaphysical puzzles.[1] But could they also teach us something about *syntax*? Well, syntactic and semantic theses can interact. And some semantic theses subsume presuppositions about the metaphysics of identity. I'll argue that when these two factors conspire, we find a syntactic puzzle in the prose of science fiction that requires the tools of both metaphysics and linguistics to solve.

Solutions to the puzzle uncover an unprecedented form of context sensitivity connected with proper names, pronouns, and descriptions. Strikingly, this sensitivity acts to manipulate even the extensions of syntactically and semantically bound pronouns. In addition to its intrinsic semantic interest, the sensitivity may have ramifications for Binding Theory, theories of persistence and, perhaps most surprisingly, the semantics of attitude reports.

1 The Puzzle

Natural languages have two kinds of pronominals: reflexives (*myself*, *herself*, *themselves*) and non-reflexives (*I*, *she*, *they*). These have systematic and cross-linguistic grammatical distribution. Some simple examples, where coindexation of subscripts, or indices, tracks coreference:

(1) (a) She$_i$ is happy.

(b) *Herself$_i$ is happy.

Draft of January 1, 2019; please don't cite without permission

✉: James.A.R.Shaw@gmail.com

† I'm grateful for helpful input from Tom Breed, Mike Caie, Ulf Hlobil, Andrei Nasta, Japa Pallikkathayil, Leona Mollica, Bryan Pickel, Katherine Ritchie, Raja Rosenhagen, and an audience at the University of Pittsburgh.

[1] See, e.g., Horwich (1975), Lewis (1976) for some classic discussions.

(c) *Jane$_i$ saw her$_i$.

(d) Jane$_i$ saw her$_j$.

(e) Jane$_i$ saw herself$_i$.

(f) *Jane$_i$ saw herself$_j$.

There are competing attempts to systematize the relevant data. But common ground among them is the idea that reflexives, like *herself*, must be syntactically bound by an antecedent. Intuitively, this means that every reflexive must have an antecedent which gives the reflexive its meaning. A more precise grip on syntactic binding isn't required for my purposes beyond the following: When x syntactically binds y, x and y meet several constraints, among which are that x and y are coindexed (tagged with the same subscripts). And coindexed referring expressions are understood to corefer.[2]

My puzzle begins by noting that in science fiction cases, the thesis that reflexives are obligatorily bound in this way seems to under-predict the number of readings of sentences like (2).

(2) Jane hits herself.

For example:

> Jane, a brilliant physicist, constructs a time machine in 2050 and travels back to the year 2016. As a young scientist in 2016, Jane is completely bewildered when a craft appears in her living room. The future traveler, highly disoriented from her trip, exits the craft in an obfuscating cloud of vapor. Confused and frightened, Jane hits herself and runs from the room.

[2] E.g., "The NP on which a reflexive is dependent for its interpretation is the antecedent... We use coindexation to indicate that [the reflexive antecedents] have the same referent." (Haegeman (1994) p.207). "NPs that are coindexed with each other are said to corefer (i.e., refer to the same entity in the world)." (Carnie (2013) p.138). The restriction to 'referring expressions' here is needed to exempt cases of semantic binding, like *Every scientist$_i$ doubted herself$_i$* where, strictly speaking, the anaphor doesn't refer at all. There are actually some further complications here that I am glossing over, but they won't matter for the aspects of Binding Theory that I mean to be tapping into. See Heim (2009) for a recent discussion of the obstacles to developing a notion of semantic 'covaluation' relevant to Binding Theory, and a proposal to overcome them.

Let me introduce two new names. Let *Future Jane* refer only to the person who exits the craft, and *Past Jane* refer only to the younger person who witnesses its arrival.[3] Then (2) seems to have four readings paraphrased below: a pair of slightly less natural symmetric readings ((3a),(3b)), and a pair of more natural asymmetric readings ((3c),(3d)).

(3) (a) Past Jane hits Past Jane.

(b) Future Jane hits Future Jane.

(c) Past Jane hits Future Jane.

(d) Future Jane hits Past Jane.

The existence of symmetric readings should be uncontroversial, up to the question of whether there are two. That latter question (of whether there are two symmetric readings) itself turns on the controversial question of whether *Jane* has multiple readings. Let me bracket that issue for later consideration.

Asymmetric readings, by contrast, are odd enough that one might suspect that they shouldn't exist. But discussions of time travel are rife with reflexives which only make sense if read asymmetrically. Some examples (my emphases) from novels ((4)-(5)), film synopses ((6)-(7)), and philosophical discussions of time travel ((8)-(9)):

(4) I don't want to die. But I've seen my own dead body. I've *seen myself* in the act of dying.[4]

(5) I don't usually *tell myself* stuff ahead of time unless it's huge, life-threatening, you know? I'm trying to live like a normal person. I don't even like *having myself around* so I try not to *drop in on myself* unless there's no choice.[5]

[3] As will become clearer by the end, to work in the way I want *Past Jane* and *Future Jane* may not properly refer to 'persons' at all. Also, strictly speaking, I don't want to presume in giving my arguments that Past Jane and Future Jane are non-identical, since some might deny that (see n.23 below). But I will treat them here as non-identical for expositional purposes. If pressed, I could make many of the same points I'm making here by replacing some of my talk about Past Jane and Future Jane with indirect talk about the spatiotemporal regions containing 'each' of them, which all should agree are non-identical. I avoid these circumlocutions here, since my aim here is just to give a rough specification of my puzzle.

[4] Gerrold (2003) p.109.

[5] Niffenegger (2003) p.21.

(6) Joe, a looper, *encounters himself* when his older self is sent back in time to be killed.[6]

(7) Aaron goes though the days several times and even *meets himself* and *gets in a fight with himself*.[7]

(8) If [a time traveler] doubles back toward the past, but not too far, he may be able to *talk to himself*.[8]

(9) Now imagine I travel back in time and *meet myself* in 1977.[9]

Reflexive constructions are completely natural, and even challenging to forego, in discussing time travel cases where a person 'encounters herself'.[10] The asymmetric uses here are too pervasive and systematic to be without explanation.

The problem is that four simple theses seem to entail that such readings don't exist. As applied to (2), they are:

(A) The only relevant interpretable constituents of (2) are *Jane*, *hits*, and *herself*.

(B) *Jane* has at most two interpretations in (2).

(C) *hits* has one interpretation in (2).

(D) *herself* in (2) corefers with *Jane*.

If (A) holds, the truth-conditions of (2) are determined by the interpretation of 3 constituents. But (B)–(D) allow for at most two permutations of the those interpretations. If (A)–(D) hold, there cannot be four truth-conditionally independent readings of (2).

The question is: Should we accept this conclusion? Or, if not, which of (A)–(D) should we reject, and why?

---

[6] Looper film Wikipedia (2015).

[7] Primer Reviews and Ratings – imdb (2015).

[8] Lewis (1976) p.147.

[9] Sider (2001) p.106.

[10] The phenomenon also seems to be cross-linguistic. Informal consultations with native speakers indicate that reflexives can get 'asymmetric' readings in German, French, Italian, Hebrew, Turkish, and Korean. (Though one native Dutch speaker reported trouble in getting the asymmetric reading, except perhaps by laying particular stress on one syllable in the polysyllabic Dutch reflexive.)

2 Denying the Ambiguity

I need to introduce some terminology for methodological purposes. I will use *ambiguous* as a shorthand for 'content-ambiguous'. That is, roughly: having several literal, unmodified semantic contents. *Ambiguous* is often used more narrowly to exclude some forms of content-ambiguity like context-sensitivity.[11] On my usage, a context-sensitive sentence like *You are hungry* is ambiguous (because context-sensitive and so content-ambiguous). But, of course, *you* is not lexically ambiguous, nor is the sentence itself syntactically ambiguous. The reason for introducing this usage is that I need terminology that separates out views that treat the several interpretations of (2) as having broadly syntactic and semantic sources, including those owing to semantically mediated context-sensitivity, from the alternatives.

After all, a tempting reaction to the puzzle is to accept (A)–(D) and insist, contrary to appearances, that (2) is not four-ways ambiguous in the foregoing broad sense. There are two ways one might do this. First, one could claim that (2) does not have four truth-conditionally independent readings, but a single reading that is 'four-ways underspecified'. That is, one could claim that our intuitions about the case are merely responsive to four different ways for a single reading of (2) to be true. A somewhat different strategy is instead to accommodate more than two truth-conditionally independent interpretations, but claim that the asymmetric readings are generated pragmatically.

Neither of these strategies pans out. Let's begin with the underspecification strategy. To clarify the idea, consider a simple illustration of underspecification: Suppose Mark is in a room with two mirrors. Then (10) is equally true if Mark sees his image in the first mirror, or in the second.

(10) Mark sees himself.

So there are two ways for (10) to be true. Indeed, there are many, many more. But this doesn't mean that (10) is several-ways ambiguous. (10) merely states a single, general condition without specifying details of how it is instantiated. (10) effectively states that Mark observes himself somehow, so its truth-conditions are insensitive as to how exactly the observing takes place.

On the underspecification strategy, (2) isn't ambiguous but merely underspecified in a similar way: (2) is underspecified between (3a)–(3d), and so true regardless of which

[11] E.g., Sennet (2015).

of them obtain, creating a misleading impression of ambiguity. The form of (2) could safely parallel that of (10), with a single univocal noun binding the anaphor, generating a single, but four-ways underspecified, reading.[12]

The underspecification strategy can (though need not) be motivated by an intriguing understanding of what happens in 'asymmetric' cases of hitting. On this understanding, there is just one person about—Jane—one of whose parts hits another. Jane counts as hitting herself when her younger part (say) is hitting her older part, just as I count as hitting myself when my palm slaps my knee. Thought the parts hitting and hit aren't identical, the persons hitting and hit are.

I suspect that there is something helpful in this construal of the case, and I'll be returning to it shortly. But even if, on some level, this is a helpful description of the time travel scenario, the underspecification strategy fails. This is because (2) fails linguistic tests for underspecification.

The most revealing failure of this kind involves negation tests. If (2) were really underspecified, its negation should negate all symmetric and asymmetric readings at once. This is because an underspecified claim is something like a disjunction between its various specifications, and to negate a disjunction is to negate each of its disjuncts. For example, returning to our earlier case, if we say *Mark doesn't see himself*, this entails both that Mark doesn't see himself in the first mirror, nor in the second, nor in any of the many other ways Mark could see himself.

We don't find a similar phenomenon with negations of sentences like (2). For example, when (the young) Jane sees the craft appear, and a version of herself step out raving about the future, we could say that Jane initially 'didn't trust herself' as a way of conveying that Past Jane didn't trust Future Jane. But this wouldn't seem to entail

[12] Parsons (2000) develops an event/state-based semantics for talk of time travelers that seems to make this prediction. Parsons is concerned to explain (among other data) why, if Socrates is a time traveler, the inference from *Socrates is sitting* (because one version sits) and *Socrates is in the marketplace* (because another version of him is there) to *Socrates is sitting in the marketplace* fails. The resolution comes in that sentences like *Socrates is sitting* involve tacit existential quantification over states—where Parsons seems to presume that any state in which a 'version' of Socrates participates counts as one where Socrates (*simpliciter*) participates. Then, whereas there may be a state of sitting in which Socrates participates and a state of being-in-the-marketplace in which Socrates participates, there may be no state of sitting-in-the-marketplace in which anyone participates, let alone Socrates. Parsons' presupposition that any state or event (say, of hitting) in which a 'version' of Socrates participates counts as one where Socrates (*simpliciter*) participates seemingly leads the reflexive constructions like (2) to have a single set of truth-conditions, satisfied whenever any version of a time traveler interacts appropriately with any other version.

that Past Jane didn't trust Past Jane. Past Jane could be as confident in herself as we like. Nor would it entail that Future Jane doesn't trust Future Jane. Nor would it entail that Future Jane doesn't trust Past Jane—indeed, Future Jane may have no reason whatsoever to doubt her past self, having once been in her very shoes. But all of these forms of distrust should be entailed by the negated reflexive construction, were it in fact underspecified.[13]

The foregoing concern counts against any view that gives (2) a single four-ways underspecified reading. There are more complex versions of the underspecification strategy. For example, one could claim that (2) has two readings, each of which is underspecified between some subset of (3a)–(3d). But this kind of view is much less plausible. It loses the intuitive motivations for the underspecification view. And, more importantly, we will still need at least one version of (2) to be underspecified between two readings, whether symmetric or asymmetric. And such views continue to make incorrect predictions about negations of (2). So we can safely set such views aside.

Let's turn to a more plausible way to safeguard (A)–(D): by maintaining that asymmetric readings are pragmatically generated from symmetric ones. On this view, the literal, unmodified content of (2) is always symmetric. It may have two symmetric readings: that Past Jane hits Past Jane, and that Future Jane hits Future Jane. Or it may have a single such reading: a univocal claim that Jane (who or whatever that may be) hits Jane. But somehow in uttering these symmetric claims, pragmatics may intervene to help us convey the truth-conditionally distinct asymmetric readings like (3c) and (3d).

(3) (c) Past Jane hits Future Jane.

(d) Future Jane hits Past Jane.

Processes of free pragmatic modulation seem suited to this work.[14] Consider by analogy two familiar examples (adapted from Nunberg (1979, 1995)): (11a), used by a

[13] The negation tests appealed to here (like all tests for underspecification) aren't always easy to apply, because data from the negated constructions can be inconstant. For example, *Mark doesn't see himself* might be used to convey that he didn't see himself in some salient mirrors, without excluding that he saw his indistinct reflection in a window. The nice point here is that *Jane didn't trust herself* never seems to have a reading where all four 'ways' for Jane to trust herself are denied. So we have as clear a verdict in this instance as we ever could that underspecification isn't at issue.

[14] My terminology here partially follows Recanati (2004, 2012), who treats *pragmatic modulation* as a general term subsuming several processes including pragmatic enrichment—the latter construed as modulation on which output meaning is 'more specific' than input meaning. For reasons of neutrality, however, I will not here follow Recanati in taking pragmatic enrichment to necessarily influence the proposition expressed by an utterance.

short order cook to convey (11b), and (12a) used to convey (12b) to a parking attendant while handing them keys.

(11) (a) The ham sandwich left without paying.

(b) The person who ordered the ham sandwich left the restaurant without paying.

(12) (a) I'm parked out back.

(b) My car is parked out back.

One reason that (11) and (12) have been of special interest to philosophers and linguists is because some aspects of their pragmatically mediated interpretation seem to be syntactically and semantically optional. For example, the 'completion' of (11a) with *The person who ordered* seems to be optional, in the sense that an utterance of (11a) may continue to express a proposition even absent a completion of this kind. This aspect of the (re)interpretation of (11) seems to contrast with what is sometimes called 'saturation', where context furnishes the values of pronominals or variables that are needed for a sentence to convey a determinate proposition. Saturation most familiarly occurs for indexical pronouns like *I* or *you*. But it may also occur for *leaves* in (11a), because there must be a place someone leaves from for the claim that they 'left' to be truth-evaluable, and so for (11a) to express a proposition.

It is controversial exactly how the 'free' processes of modulation in cases like (11) or (12) take place. But we can skirt the details here. What matters for us is the characteristically optional character of modulation. For it is precisely this optional character that would suit it to account for the readings of (2).

To see how such an explanation might go, let me begin by focusing on a view that takes (2) to have a pair of truth-conditionally independent symmetric readings—equivalent to (3a) and (3b)—owing to an ambiguity in *Jane*.

(2) Jane hits herself.

(3) (a) Past Jane hits Past Jane.

(b) Future Jane hits Future Jane.

Processes of pragmatic modulation could take us from these symmetric readings to the asymmetric readings in any number of ways, depending on our theory of modulation.

Perhaps we engage in 'expansion' that takes us to new syntactic structures in interpretation, for example from (2) (on its interpretation in (2a)) to (13).

(13) [Past] Jane hits the future version of herself.

Perhaps we 'transfer' the meanings of some of the terms in (2)—for example, interpreting *hits* as *hits a past version of*.[15] And there may be other accounts yet.

As will become clear soon, I sympathize with the idea that context plays a role in generating asymmetric readings. But there is an important bit of data that seems to show that the role of context is not manifested in a process of pragmatic modulation, no matter how we understand that process. The issue is that replacement of the referring expressions in (2) with unambiguous, supposedly coreferring expressions blocks the availability of asymmetric readings. But there is no reason why such replacement should interfere with a hypothesized process of modulation.

Consider again (11a). As it happens, other ways of referring to the ham sandwich in (11a) seem to preserve the possibility of the modulation that leads to (11b). These include swapping the description for a coreferring description, a demonstrative, a pronoun, or a name.[16]

(14) (a) The half-eaten mess left without paying.

(b) That one [pointing to the ham sandwich] left without paying.

(c) Check on the ham sandwich. It may have left without paying.

(a) Croque-Monsieur left without paying.

We needn't go so far as to claim that substitution of any coreferring expressions would preserve the possibility of modulation. Linguistic context may matter, and some expressions may be more conducive to modulation than others. But the data at least supports the claim that we should generally expect substitution of coreferring terms to preserve the possibility of modulation barring special explanation.

[15] See Nunberg (1995) for a discussion of meaning transfer.

[16] In the case of anaphora, the possibility of modulation seems dependent on the possibility of the antecedent being modulated. Also, intuitions about the felicity of (14c) may vary: some may feel that the pronoun should agree in gender with that of the person, not the sandwich. Neither of these complications matters for the points I'm raising here about (2).

The problem now is that the substitution of allegedly coreferring terms in (2) completely blocks the availability of asymmetric readings. For example, (3a) and (3b) simply have no asymmetric readings at all.[17]

(3) (a) Past Jane hits Past Jane.

(b) Future Jane hits Future Jane.

Recall that on the modulation strategy presently under consideration, (2) is ambiguous between two truth-conditionally independent symmetric readings, one of which is truth-conditionally equivalent to (3a) and the other of which is truth-conditionally equivalent to (3b). The hypothesis is that one of these two readings, or both, generate asymmetric readings via pragmatic processes of modulation. Accordingly, in disambiguating (2) by substituting appropriate coreferring terms for its name and anaphor, for example as in (3a), we would expect not to disrupt the availability of some modulated readings. Instead, we find that the asymmetric readings simply disappear.

There doesn't seem to be any obvious explanation of why substitution of coreferents should lead to such a striking contrast in the availability of asymmetric interpretations. I conceded that we should be open to the idea that linguistic context might sometimes influence the availability of modulation, by making it slightly more challenging or slightly easier. But all we've done in these cases is replaced a noun and a reflexive with proper names (allegedly) bearing identical denotations. In general this doesn't seem to differentially influence modulation significantly, let alone in the extreme ways we would have to posit for (2) and (3a). So this version of the modulation strategy uncovers an important disanalogy between cases like (11)–(12) and that of (2)—one that reveals we shouldn't be treating them in the same way.

The foregoing argument against modulation targeted a view that began with the assumption that (2) had a pair of literal, unmodulated readings given by (3a) and (3b). What of the view that claims that (2) is univocal, with modulation operating on that single reading? This view isn't necessarily subject to the objection just given from substitution (since not all such views will allow that *Jane* corefers with either *Past Jane* or *Future Jane* on a given reading of (2)). But it is subject to the earlier objection concerning the negations of (2). Treating (2) as univocal predicts that a sentence like *Jane*

[17] If there are concerns about whether we can have intuitions about my newly introduced names like *Past Jane*, note that uses of (otherwise idiomatic) descriptions like *the past version of Jane hit the past version of Jane* don't seem to have asymmetric readings either.

doesn't trust herself has a (literal, unmodulated) reading which negates at least a pair of readings from among (3a)–(3d). But there is no such reading. So, ultimately, no version of the modulation strategy is viable.

Having rejected the utility of pragmatic modulation, let me quickly dispense with a different pragmatic explanation that may have seemed attractive: an appeal to conversational implicature. One could claim that (2) literally asserts only symmetric readings, and asymmetric readings are merely implicated. This is the least plausible view considered so far. First, the standard mechanics for the generation of a conversational implicature involve asserting something that violates some norm of truth, relevance, politeness, and so on. An immediate problem is that we can obtain readings of (2) that are asymmetric, even if the corresponding symmetric version is true, relevant, polite, suitably informative—virtuous along almost any dimension one could think of. Second, we still need an account of why (2) generates the implicature, whereas (3a) and (3b) do not, since conversational implicatures are supposed to be nondetachable. Again, if we're told that Past Jane trusts Past Jane, it is almost impossible to hear this as implicating that Past Jane trusts Future Jane. Implicature seems no better placed than free pragmatic modulation to make the transition here. Indeed, if anything, it is worse, since it is especially mysterious what reasoning would lead us from symmetric claims to implicated asymmetric claims.

3 Tinkering with Syntax

So far, I've argued that we should take appearances at face value: (2) really does have four truth-conditionally independent readings (not merely one or two underspecified readings). And (2) does not convey the asymmetric readings through pragmatic modulation or implicature. Instead, it does so as a result of its literal, unmodulated semantic makeup. As a result, we are again confronted with the hard choice of denying one of (A)–(D).

(A) The only relevant interpretable constituents of (2) are *Jane*, *hits*, and *herself*.

(B) *Jane* has at most two relevant interpretations in (2).

(C) *hits* has one relevant interpretation in (2).

(D) *herself* in (2) corefers with *Jane*.

Denying (B) is a non-starter. Even if *Jane* has more than two readings, it's unclear how these additional readings, on their own, get us any closer to generating asymmetric readings. So we're left with (A), (C), and (D).

In this section, I want to consider an attempt to resolve difficulties by denying (A). On the suggested view, phonetically unrealized interpretive material relevant to symmetric readings may reside in the syntactic structure of (2). As will become clearer soon, I am open to this as a strategy for coping with the puzzle. But for now I will continue to proceed negatively, by arguing against an oversimplified version of this strategy: one which enriches the syntactic structure of (A) to such a degree that we distort its argument structure, so that *Jane* ceases to be the subject of the verb *hits*, or *herself* ceases to be its object.

This strategy may seem attractive if we compare the asymmetric readings of (2) to the sentences in (15).

(15) (a) Jane$_i$ hits [the future version of [herself]$_i$]$_j$

(b) Jane$_i$ hits [the past version of [herself]$_i$]$_j$

In obtaining the asymmetric claims in (15), no special problem is posed to our theories of syntactic binding. The object of *hits* isn't supplied by the reflexive, but by an expression containing the reflexive as a constituent. The denotation of the embedding description can vary independently of that of the antecedent *Jane* if desired, as indicated by the different indices. So if there were a process whereby (2) were ambiguous between two syntactic structures something like (15a) and (15b), with the relevant descriptive material being phonetically unrealized, we would have made progress on our puzzle.

The problem is that this view takes on commitments about the argument structure of (2) that aren't borne out. Since we are positing a different object of the transitive verb than the reflexive, we should see the reflexive exhibit shifts in case depending on the choice of verb and embedding null material. But we don't.

To see this, we need to consider a language other than English, since English reflexives lack relevant case-marking. But the first person reflexive in German, for example, has case-marking that allows us to test the current view. *Mich*/*mir* has different case (accusative and dative respectively) in (16a) and (16b) depending on the absence or presence of the embedding descriptive material. Both, German consultants claim, have asymmetric readings in time travel cases. But (16c), which lacks the descriptive material but retains the case-marking of (16b), is infelicitous (except in special dialects)—ostensibly because it can't be read as containing the covert material.

(16) (a) *Ich schlage mich.*
I hit myself-ACC.
'I hit myself'

(b) *Ich schlage die zukünftige Version von mir.*
I hit the future version of myself-DAT.
'I hit the future version of myself'

(c) #*Ich schlage mir.*
I hit myself-DAT.
'I hit myself'

Now, there are of course several different possible versions of the view I'm criticizing, which differ in their accounts of what material embeds the reflexive *herself* in asymmetric readings of (2). The embedding material needn't be a description, for example. But any version of the view that significantly alters argument structure will face versions of this same problem, precisely because the account will crucially turn on the availability of covert material to supplant *herself* as the direct object of *hits*. For any proposal that does this, we can select a transitive verb that should create a case mismatch between the verb's object and the reflexive. In other words, regardless of what the embedding material is proposed to be, we can create problematic examples like (16a)–(16c). So treating this kind of embedding material as a covert part of syntactic structure is untenable.[18]

Before moving on, I want to comment on a phenomenon which may seem to improve the prospects of using covert syntactic material to explain the asymmetric readings. When I considered possible reinterpretations of (2), I focused on elaborations like (15a) and (15b).

(15) (a) Jane$_i$ hits [the future version of [herself]$_i$]$_j$

(b) Jane$_i$ hits [the past version of [herself]$_i$]$_j$

The problem for treating added syntactic material as covert was that it implausibly supplanted *herself* as the object of *hits*. But there are other idiomatic constructions which may appear to avoid this consequence. For example:

(17) (a) Jane hits her future self.

[18] This view may also face the difficulties in explaining the loss of asymmetric readings on the substitution of coreferring terms, discussed for the pragmatic modulation strategy in §2.

(b) Jane hits her past self.

Talk of one's past and future selves is a natural way of generating asymmetric claims in time travel cases. Consider an example cited above (altered emphasis mine).

(6) Joe, a looper, encounters himself when *his older self* is sent back in time to be killed.

Appealing to *her past self* could seem significantly different from appealing to *the past version of herself*, because *her past self* seems like a special kind of compound reflexive—a construction in which a reflexive is disambiguated by making it more specific. We might consider *her past self* a 'complex reflexive' by analogy with a complex demonstrative like *that dog in the window*. And one could hope that appeal to such constructions would strengthen the case for denying (A) that I've been criticizing by avoiding disruptions of argument structure.

But this view is much less plausible than it may initially seem. There is a temptation to think that to get from (2) to (17b), we need only insert the adjective *past* between the components of a morphologically complex reflexive as in (18).

(18) Jane hits her[-past-]self.

While it is true that the English reflexive may be morphologically complex, it does not have components corresponding to the relevant components of (17b): the possessive *her* and the noun *self*. It is an idiosyncratic feature of English that the root and prefix of a morphologically complex reflexive correspond to the possessive and noun. Languages with monomorphemic reflexives like German or French reveal that the strategy of appealing to 'complex reflexives' cannot properly generalize. It is hopeless, for example, to claim that the asymmetric readings of the German (19a) owe to the presence of syntactic material that 'enriches' the reflexive *sich* along the lines of (17b). Not only is *sich* morphologically simple, but its morphology is simply absent from (19b)—the relevant translation of (17b).

(19) (a) *Jane schlägt sich.*
Jane hits herself.
'Jane hits herself'

(b) *Jane schlägt ihr früheres Selbst.*
Jane hits her earlier self.
'Jane hits her earlier self'

Sentences like (17a)–(17b) are, I think, of independent interest. The problem is that ascertaining how they are able to convey asymmetric readings cannot not help us with our problems with (2). (17a)–(17b) simply don't contain reflexives, and it is the presence of the reflexive in (2) that makes our puzzle hard to solve.

All this is showing us that if we want to enrich the syntactic structure of (2) to cope with our problems, we will have to do so more carefully, and in ways that substantially improve on the two strategies considered so far.

4 Metaphysical Assists?

Underspecification, covert material, free pragmatic modulation, and implicature have all proven to be of little help. Some may be concerned that these dead ends indicate that I've been looking in the wrong place for a resolution to the puzzle. My approaches so far have made 'mere' appeals to linguistic mechanisms, considered in abstraction from the metaphysics—for example the ontology—operative in the time travel case. As I noted at the outset, time travelers pose all sorts of conundrums about existence and identity. Maybe a more metaphysical approach will yield greater insight.

I think this is a step in the right direction. Nevertheless, I want to argue here that getting clear on the metaphysics *alone* won't suffice to resolve our puzzle. There are all sorts of metaphysical debates germane to time travel: debates about the metaphysical or physical possibility of time travel, debates between the presentist and the eternalist about whether the future and the past exist, debates between the three- and four-dimensionalist about whether or not objects are 'spread out' in time, debates between the perdurantist and the endurantist about how objects persist through time, and debates about what grounds personal and objectual identity over time. But the first debates (about the metaphysical or physical possibility of time travel) are irrelevant to my puzzles as long as time travel can be epistemically possible (which it can). Moreover, my cases involve objects, acts, and events that all exist or take place at one time: 'the present'. As such, (2) raises a puzzle whose formulation and force largely skirt the remaining debates in the metaphysics of time.

To see why this latter claim holds, let me introduce a little terminology. I'll speak of the referents of ordinary language referring expressions like *Jane* as *continuants*.[19] Call the time at which the hitting takes place *t*. While the hitting transpires at *t*, there is a

[19] Cf. Sider (2001) p.60, though this particular usage of *continuant* is not universal.

youngish humanoid shape (covering two arms and two legs, for example). Let's denote by "*A*" the mereological sum of everything existing at *t* subsumed within that shape. There is also an older humanoid shape. Let's denote by "*B*" the mereological sum of everything existing at *t* subsumed within that shape.

We can group views about continuants into two classes, depending on how the referent of *Jane* (at a context) relates to *A* and *B*.

> *Type I*: the referent of *Jane* (in context) contains *A*, but not *B*, as a part, or vice versa.

> *Type II*: the referent of *Jane* (in context) contains both *A* and *B* parts.

Note that these views are defined on the basis of whether they take *A* or *B* to be parts—not necessarily proper parts—of the referent of *Jane* (in context). A Type I view might take the referent of *Jane* to have *A* as one of its spatiotemporal parts. But it might also take the referent of *Jane* to simply be identical with *A*.

Relatedly, note that this classification cross-cuts important metaphysical disagreements. For example, a three-dimensionalist who takes *Jane* (in context) to refer to one of two two-legged, two-armed, 'wholly present' beings at *t*, and a four-dimensionalist who takes *Jane* to refer to a two-legged, two-armed, stage of a four-dimensional spacetime worm, will both be theorists grouped under the Type I heading.[20] Moreover, a four-dimensionalist who takes *Jane* to refer to a spacetime worm that subsumes both her younger and older selves at *t*, and a three-dimensionalist who takes *Jane* to refer to a single 'wholly present' bi-located entity, will both count as Type II theorists.

Views about continuants are semantic-cum-metaphysical views. They pronounce not only on what things exist, and on what they are like, but also on which of them we ordinarily talk about. As such, these views are to some extent hostage to semantic data. Here, then, is the problem: Type II views, without supplementation by significant and controversial linguistic theses, are simply implausible. But Type I views (again without such supplementation) don't resolve the puzzle about binding raised by (2).

[20] In discussing the three- and four-dimensionalist views, I'm bracketing questions about their compatibility with time travel. See Lewis (1976) for arguments that four-dimensionalism is compatible with time travel, Sider (2001) §4.7.2 and Simon (2005) for arguments that three-dimensionalism is not, Markosian (2004) and Miller (2006) for replies on behalf of the three-dimensionalist, and Gilmore (2007) for an argument that time travel presents challenges even for the four-dimensionalist.

To see why linguistically unsupplemented Type II views are implausible, contrast some different ways of talking about Jane. Some of these ways of speaking seem to yield truths, or at least true readings, evaluated at *t*.

(20) Jane is a woman.

(21) Jane is a scientist.

(20) and (21) aren't completely unproblematic. But they don't sound immediately objectionable or confused even in the context of the time travel story. Indeed, in time travel fiction where an individual 'overlaps' with themselves at a particular time, one finds a good deal of discourse about that time traveler proceeding roughly along the lines of (20)–(21).[21]

Contrast these with claims that should be unambiguously true at *t* if the relevant Type II views are right:

(22) Jane has [at least] four arms and four legs.

(23) Jane has a few square feet of empty space in her middle.

(22) should be true because any entity that subsumes both *A* and *B* is an entity that has at least four arms and four legs. (23) should be true because any entity that subsumes both *A* and *B* also has a large gap in its middle region. But there do not seem to be intuitively true readings of (22) or (23). Or, at least, one does not find ordinary speakers making such pronouncements when discussing cases of time travel.[22] Since Type II

[21] Cf. the cases discussed in Parsons (2000) (and n.12 above).

[22] I want to grant that there may be some special uses of *Jane*, and of descriptions or demonstratives, that are more likely to accord with something like a Type II view. It might be correct, faced with Past Jane and Future Jane, to report that the woman on the left is the woman on the right, seemingly employing the *is* of identity (as a way of conveying that the two individuals in front of one are versions of a 'single' time traveler). Even if there are such uses, this won't detract from the point that no instances of (22) or (23) seem assertable. If we can make the relevant identity claims, we probably need our semantics to accommodate contextual shifts to an 'atemporal' or 'timeless' perspective (if we are four-dimensionalists) or a perspective of 'coarse-grained individuation' of persons (if we are three-dimensionalists) to capture those readings. (Indeed, we may need something even more sophisticated anyway if we adopt something like a four-dimensionalist view on which we typically refer to something like person-stages. Such a view probably needs a way of picking out entities that span a temporal interval determined by context (cf. Sider (1996) pp.448–9 and especially Moss (2012) pp.674*ff.*).) All views I will eventually sketch in §6 should be able to accommodate shifts between discourse about (say) whole spacetime worms and their spatiotemporal parts either as-is, or with minor modifications. The

views come with commitments about the semantics of ordinary English discourse, they make erroneous predictions about the behavior of simple predications, at least to the extent they aren't supplemented with additional controversial linguistic theses.[23] The data even rules out a view, which I didn't list above, on which *Jane* is pervasively context-sensitive between a Type II reading and one or more Type I readings.

This means that the only plausible treatment of continuants, in the absence of linguistic maneuvering, would require a Type I semantics and metaphysics. But as soon as we adopt such a view, the puzzles about (2) are as forceful as ever. Any reading of (2) will be one on which *Jane* refers to an entity subsuming *A* but not *B*, or *B* but not *A*. Constraints on binding will then ensure the anaphor will share this denotation. This seems to ensure we have a symmetric reading. But the puzzle is precisely to say how asymmetric readings are obtained. Getting clearer on the metaphysics just seems to reinforce the problems.

As already noted, I think that reflecting on the metaphysics of my cases is a step in the right direction. But thinking about the metaphysics alone doesn't get us very far. If metaphysics can help provide a resolution of our puzzle, it will have to interact with

---

important point is that such alternative perspectives are not operative in ordinary predications as in (22) or (23), hence not in our predication in (2). When Jane hits herself, it is with one of her two arms, not one of her four. To rule out the (unsupplemented) Type II semantics for (22) and (23) is to do so for (2).

[23] Of course, this objection against Type II views will not longer have force if we endorse some additional interesting linguistic (or linguistic-cum-metaphysical) theses that explain why (22) or (23) come out false. This is fine, since my argument here is merely that metaphysics *alone* won't resolve our puzzle. Still, let me mention two linguistically revisionary strategies to salvage Type II views, by way of previewing. The first is to take the truth-conditions of a sentence like *Jane sits* to be individuated more finely than by the referent of *Jane* and the property of sitting. For example, we could take a four-dimensionalist view on which *Jane* refers to a spacetime worm subsuming *A* and *B*, but have the truth-conditions of *Jane sits* somehow turn of how things stand with a spatiotemporal part of that spacetime worm like *A* (e.g., with covert counterpart functions). Another view might take *Jane sits* to be evaluated on the basis of 'relativized' properties. For example, a three-dimensionalist might take *A* and *B* to be identical, with *Jane* referring to both, and then relativize properties or property instantiation to spatial location, on the model of relativizations of properties to times (Van Inwagen (1990)) or property instantiation to times (Johnston (1987), Lowe (1988), Haslanger (1989)). Indeed, just such a position has been suggested by Sider (2001) as a possible best option for the three-dimensionalist to try to accommodate time travel, and is discussed and defended in Miller (2006). I take no stand on whether any of these kinds of views is metaphysically defensible: my primary concern is to capture the semantics of ordinary discourse. Each of the views just described needs significant and controversial semantic maneuvers to account for such discourse, and I'll consider both from a more linguistically-minded standpoint shortly: a view that relativizes properties in §5, and a broadly counterpart-theoretic treatment in §6.

linguistic theses to do so. So let's return to consider our remaining semantic options.

5 Semantic Indeterminacy and Contextual Resolution

We need an explanation of how (2) has asymmetric interpretations with semantic, compositional sources. And this explanation should keep (2)'s syntax—for example, its argument structure and the presence of an anaphor—relatively close to its surface grammar.

In this section, I'll work my way up to two proposals that can accommodate the data. And to do this, I want to begin by focusing on a simpler aspect of the semantics of (2): the apparent ambiguity of its proper name. After surveying some data relevant to that ambiguity, I'll give a very broad theoretical explanation of its source. And with that explanation in hand, I'll detail some more specific implementation strategies that cover the data from reflexives, but between which I'll remain neutral.

Let's start with three pieces of evidence that names, even in simple predications like *Jane sits*, exhibit noteworthy (content) ambiguities. First, and most obviously, this hypothesis would explain the apparent availability of interpretive shifts in descriptions of time travel. Simple predications can apparently facilitate talk about different 'versions' of an individual based on appropriate contextual cues. For example, in the story fragment below, it's easy to read the first pair of *Jane*s as concerning Future Jane, and the latter pair as concerning Past Jane.

> Jane arrived from the future in her craft. Jane stepped out and looked around at her old house and wandered into her old room. In a bed in that room, Jane was still sleeping, unaware of the time traveler's presence. But Jane would soon take part in an adventure she never could have anticipated.

Just as time travel stories are replete with asymmetric reflexives, they are also replete with uses of names like those above, where context helps resolve which of two versions of a time traveler we are talking about.

Second, the hypothesis explains an important aspect of the data considered already in discussions of the pragmatic modulation and syntactic enrichment strategies. If there is no ambiguity associated with the sentence fragment *Jane hits* _____, then we should predict that in (2) and similar reflexive constructions no ambiguities trace to the subject or verb. We could perhaps put the point most forcefully in terms of predictions

about negations. If *Jane hits* introduced no ambiguity, there should be a single symmetric reading of (2) and other relevant reflexive constructions. But that means that there should be a single 'negated' symmetric reading, and that is not true. For example, the symmetric reading of *Jane doesn't trust herself* can convey that Past Jane lacks self-confidence, without claiming that Future Jane does. Allowing the initial predication to introduce ambiguity in the sentence generates the correct number of symmetric, and negated-symmetric, readings.

Third, the hypothesis gives a nice explanation of judgments governing variations on time travel cases in which we add and subtract 'overlap'. Suppose Past Jane is sitting at *t*, and Future Jane is standing, hence not sitting. Suppose someone were to ask: Is Jane, at *t*, sitting or not? It isn't obvious, I think, how to answer this question with a simple *she is* or *she isn't*. Suppose, however, that Jane never travels back in time at all. At *t* there is only Past Jane sitting. If asked: Is Jane sitting at *t*, or not? The answer is: She is. Suppose instead that Jane were born many years after *t*, but travelled back to that time, before her birth. At *t* there is only Future Jane, standing and gazing at her surroundings after exiting her craft. Is Jane sitting at *t*, or not? Surely not.

The pattern of reactions here is what one would expect if *Jane is sitting* not only introduced an ambiguity in the 'overlap' case, but an ambiguity between something like Past Jane and Future Jane in particular. This would explain why the relevant predication can seemingly characterize either Past Jane when Future Jane is taken out of the picture, or Future Jane when Past Jane is removed. But when both characters are present, and in a contextual vacuum, the question of whether Jane is standing or not is seemingly best addressed by asking for clarification: which version of Jane are you talking about?[24]

Note that the second point above is a small piece of evidence that we have not merely a plurality of interpretations of *Jane sits* of some kind or other, but a form of what I've called 'content ambiguity' in particular. And more generally, we clearly should combine the three foregoing points with the data from reflexive constructions like (2). It should be clear that what we have on our hands is a single phenomenon that calls out for a unified explanation. This means, for example, that we should avoid understanding the ambiguity in *Jane sits* as owing to pragmatic modulation or implicature. Nor should we

[24] Note that the claim that I am arguing for here is merely that simple predications can introduce ambiguity into a sentence in cases of overlap, not that they must. Indeed, when we talk about 'versions of Jane', as I just have, it seems that we use *Jane* univocally, so that both the younger and older women count as versions of one person. Cf. the remarks on talk from a 'timeless' or 'coarse grained' perspective in n.22.

rely on syntactic enrichment that would disrupt argument structure: such an account is in principle unable to generalize to explain the behavior of reflexives.

This leaves us with two questions: What is the syntactic source of the ambiguity? And what kind of ambiguity is it? Let's begin with the first question.

There are views that might seek to connect the ambiguity of *Jane sits* with its verb. Consider one such view, framed against the backdrop of a four-dimensionalist metaphysics: We let *Jane* refer always and only to a single spatiotemporally extended spacetime worm. Then, during a time travel overlap at t, we associate *sits* with multiple possible interpretations that differ in their extensions at t (this could, for example, be a form of lexical ambiguity, or of context-sensitivity). According to the first interpretation, a spacetime worm 'sits' at t if it has a 'younger part' bent in the appropriate way at t.[25] And according to the second interpretation, a spacetime worm 'sits' (at t) if it has an 'older part' bent in the appropriate way at t.[26]

Any view which traces the ambiguity to the verb in this way creates an unintuitive proliferation of properties. This is clearest when we consider that a time traveler can 'revisit' a single time more than once. As revisitations increase, ambiguities grow linearly for simple predications, and quadractically for transitive reflexive constructions (and cubically for ditransitives). Intuitively revisitations multiply (if anything) agents, not property or relation types. The proliferation of properties leads to troubles for (linguistic) coordination. Jane may visit the past with her assistant Mary. At some point, the older version of Mary and the younger version of Jane separate from the group, and become lost. The older version of Jane recruits Mary's younger self in a search for them. We can report that they each succeed in finding their corresponding counterparts with (24).

(24) Each of Jane and Mary eventually found herself.

If ambiguities trace to the verb, we somehow need *found* to disambiguate in two different ways at the same time. We could conceivably continue to proliferate interpretations

[25] Where the relevant 'parts' should be understood as something like time-slices, sliced along personal time (see Lewis (1976) for a discussion of personal time). Cf. Sider (2001) p.101.

[26] Cf. Lewis' treatment of temporary intrinsics as properties perduring four-dimensional objects bear in virtue of intrinsic properties had by their spatiotemporal parts (Lewis (1986) p.204). For another view that might try to locate ambiguities in predicates, see the citations in n.23 regarding the three-dimensionalist view that treats time travelers as 'wholly present' in multiple locations at once and relativizes properties (or property instantiation) to location.

of the verb to avoid the problem, but the view is starting to become unwieldy, in addition to looking *ad hoc*.

So it is preferable to connect the source of the plurality of interpretations to the relevant nouns, like *Jane*. This leaves our last question: what kind of ambiguity is it? We can helpfully rule out a lexical ambiguity. It would obviously be problematic to take an oversimplified view on which there were an indefinite range of *Jane*s that we had to learn separately to talk about the single time traveler. The more natural approach to generate lexical ambiguity would allow for a single learned entry (per name), which could generate a plurality of additional entries via a single overarching lexical rule. But even this approach is unhelpful, because the strategy cannot extend to cover reflexive pronouns. It is not clear what the analog of the relevant lexical rule would be for pronouns, nor why it would be in any way helpful to have multiple lexical entries corresponding to *herself*—that wouldn't get us any further in settling how we get asymmetric readings in cases like (2).[27] Since we've already ruled out pragmatic forms of ambiguity, syntactic ambiguity, and lexical ambiguity, this leaves one salient option by a process of elimination: context-sensitivity.

We've now finally been able to narrow down the source of the ambiguity in our reflexive constructions a great deal: *it must arise from a form of context-sensitivity connected with the (unmodulated) semantics of referring terms*. Before discussing two ways to implement this, it will be important to step back and gain a broader theoretical understanding of why exactly this context sensitivity is arising. The moral is challenging to state in theoretically neutral language. But let me try to do so, as best as I can, for names.

Let's focus on what I'll call an 'ordinary proper name' like *Jane*. (Examples of proper names that are not 'ordinary' are those I introduced in §1, like *Past Jane*.) In gaining competence with such a name, one must come to appreciate conditions for reapplying the name to individuals at alternate times and worlds. In doing so, one leans on an appreciation of conditions under which some object o that exists in a world w and a time t can count as the extension of the name at that world and time. I don't mean to

[27] Note, I of course want to allow that *herself* could be associated on different uses with distinct variables in syntax. The trouble is of course that each of these variables would seem to play the same semantic role as any other unless, as I consider shortly, we do something like relax constraints on the semantic 'covaluation' connected with coindexation. There are complex, troubling technical and philosophical questions in the vicinity here (see especially Fine (2007)). But I don't think they directly pertain to the issues I'm addressing here, so I set them aside.

presume that such conditions are a component of the semantic value of a name. They may draw predominantly on metasemantic information. And I don't want to presume any one story about what such conditions look like. To give an idea of what I have in mind: Perhaps for an object *o* existing in *w* at *t* to stand as the extension of *Jane* is for *o* to be numerically identical with an object *o* that was baptized in the actual world with that name. Or, to take a different account, perhaps for some *o* existing in the actual world at *t* to stand as the extension of *Jane* requires *o* to be a time-slice 'counterpart' of a time-slice, originally and actually baptized *Jane*, under suitable relations (say, something like psychological and causal continuity). Or perhaps the conditions do not rely on a notion of baptism at all. The important point is that there be some such conditions, which I will call the *referential conditions for a name N relative to w and t*.[28]

Time travel cases with 'overlap' generate numerically distinct objects that are equally good candidates for meeting the referential conditions of a single name. Indeed, each of several objects would individually seem to uniquely meet those conditions, were it not for the presence of the other objects.[29] This is precisely what we see occur for cases where we add and subtract time travel 'overlap': if only Past Jane is about, *Jane* unambiguously takes her as an extension. Whenever two or more objects each have this standing with respect to an ordinary proper name, I will say that they have *equal ordinary referential standing*.[30]

[28] I've spoken here of the conditions that allow an object to be the extension of a proper name. But if the logical form of our sentences turns out to have hidden elements that intuitively 'shift' that extension—e.g., hidden counterpart functions (cf. the view sketched near the end of §6)—then it is the object 'post shift' that we are interested in. As such, the ensuing definitions may have to be adjusted to reflect this. Here, I'll stick to the simpler formulations.

[29] A caveat: we may not be able to frame things in this way if the objects in question are, say, world-bound time-slices. I hope the reader won't begrudge me a pinch of salt. Also, note that depending on what referential conditions for a name are, we may end up describing this 'competition' between referential candidates in different ways. Perhaps we will say that there is indeterminacy in which meets the conditions. Or perhaps we will say that (determinately) several objects completely meet the conditions. Again, I mean the current formulation to be neutral between these options.

[30] As currently formulated, this definition is non-ideal since it seems to make equal ordinary referential standing parasitic on which names are present in language. We can expand the definition to include objects that 'would' have this status with respect to an ordinary name. A deeper concern for the foregoing definitional strategy is that it makes equal ordinary referential standing parasitic on something like naming conventions. But this is not obviously undesirable. We are investigating the semantics governing continuants: those objects that figure as the ordinary referents of names, and populate ordinary quantifier domains. There are good reasons to think that which kinds of objects play this role (and, in particular, what the persistence conditions and modal properties of these objects are) reflect some degree of conventionality. It may not be possible to characterize the notion we are after without

What seems to happen, when we use a name like *Jane* that seeks an extension at a world and time among several objects that have equal ordinary referential standing, is that context may intervene to settle which of the several objects the speaker is talking about. As I'm envisaging it, this role for context is as sensitive to our idiosyncratic interests as the involvement of context in settling the bounds of quantifier domains. Accordingly, I won't try to say what the conditions are for context to pick out one object rather than another in this process of disambiguation. I doubt there is any informative general characterization of these conditions.

What (2) shows is that this kind of context-sensitivity needs to be extended from names to syntactically bound pronouns. Indeed, as already seen above, it should also be extended to semantically bound pronouns. We can get a natural asymmetric reading of a sentence like (25).

(25) Every time traveler trusted herself.

As one might expect, the sensitivity also seems to extend to deictic pronouns and (though it may be trickier to get the relevant readings) definite descriptions. For example, Future Jane and Past Mary may be out looking for Past Jane, when Past Mary sees her in the distance at the edge of a lake. Past Mary may draw this fact to the attention of Future Jane, addressing the report in (26) to her.

(26) You're standing by the edge of the lake.

Intriguingly, (26) seems to admit such an interpretation even though Past Jane, who is standing by the edge of the lake, is not the addressee.[31] A parallel phenomenon may occur with definite descriptions, though it may require very special contexts. Suppose you and I are assisting Future Jane and Past Mary look for Past Jane. Assume that all versions of Mary and Jane are scientists. I have forgotten who we are looking for and ask you which of these two time traveling companions we're seeking again. With Future Jane standing to the left of Past Mary, you could reply with (27).

(27) We're looking for the scientist on the left.

The relevant reading of (27) may be harder to arrive at than the others I've discussed. But I won't dwell on why. Instead, it's time to start exploring two broad strategies for implementing the kind of context-sensitivity I've been describing.

---

appealing to something like the semantics of 'ordinary names' or 'ordinarily used quantifiers'.

[31] Thanks to [acknowledgement omitted for blind review] for pointing out a similar example.

6 Two Semantic Implementations

I framed the puzzle concerning (2) as requiring us to deny one of (A)–(D).

(A) The only relevant interpretable constituents of (2) are *Jane*, *hits*, and *herself*.

(B) *Jane* has at most two relevant interpretations in (2).

(C) *hits* has one relevant interpretation in (2).

(D) *herself* in (2) corefers with *Jane*.

I ruled out denying (B) as unhelpful, and in the previous section argued that (C) comes with high theoretical costs. That leaves (A) and (D). Here, I'll explore how we might get around the puzzle by denying each of these theses, starting with (D). Those who are uninterested in the gritty details can skip to the final section without too much loss.

I'll begin by helping myself to a slightly amended version of an individual concept (though I will continue to use that terminology). Ordinarily, an individual concept is a function from worlds or world-time pairs to an object. My expanded notion of an individual concept will map world-time pairs to sets of objects. Intuitively, an individual concept in my sense is a function that, for some possible 'ordinary' name, maps a world-time pair to the objects at that world and time that are possible extension assignments for the name. As such, my individual concepts will ordinarily map a world-time pair to a singleton, consisting of some lone object, existing at the relevant world and time, that would meet the referential conditions for the hypothetical name. Perhaps the function sometimes maps a world-time pair to the empty set (when no such object exists). But in extraordinary cases (such as time travel overlap) it may return a set containing several objects with what I've called 'equal ordinary referential standing'. I'll denote the individual concept associated with an actual ordinary name, like *Jane*, with small caps: JANE. And I'll assume, perhaps oversimplifying, that the set of all ('ordinary') individuals that exist at a world and time are partitioned by the ranges of all individual concepts: every individual that exists at world w and time t is a member of the output of exactly one individual concept. For example, if t is a time in w at which the only time travelers are Past Jane and Future Jane, then every individual other than Past Jane and Future Jane is associated with an individual concept that maps $\langle w, t\rangle$ to the singleton containing them, and Past Jane and Future Jane are associated only with the single individual concept JANE such that $\text{JANE}(\langle w, t\rangle) = \{\text{Past Jane, Future Jane}\}$.

Next, I will assume that a context c provides a selection function S_c, which is a (perhaps partial) function that maps a set of individuals to one of the individuals in the set. The idea is that these selection functions will operate on the set values of the aforementioned individual concepts at world-time pairs. When an individual concept returns a singleton (as will occur in all ordinary cases), the contextually determined selection function has no work to do besides mapping the singleton to its lone member. The real interest of the selection functions is in their ability to pick out a contextually distinguished candidate from non-singleton sets like $\textsc{jane}(\langle w, t\rangle)$. When the individual concept returns a non-singleton set, context must select among its members for compositional processes to engage. We can suppose that in defective contexts, the selection function is undefined for these non-singleton sets.

On the first implementation I'm now exploring, we deal with simple predications by making proper names like *Jane* context-sensitive. The name's extension at an index-context-assignment triple is given by applying the contextual selection function to the value of the name's associated individual concept at the index.

$$\text{Jane}^{w,t,c,g} = S_c(\textsc{jane}(\langle w, t\rangle))$$

This gives *Jane sits* two interpretations (relative to the pertinent world-time pairs) depending on context, as desired, without any further complications.

Dealing with pronouns, and syntactic and semantic binding, will be a trickier affair. Let's start with deictic pronouns.

Here I'll work with a treatment on which pronouns are tagged with numerical indices and operate as variables that get their values from an assignment function g that maps numerical indices to individuals. But to accommodate 'shifted' readings of pronouns, I'm going to assume that context supplies a counterpart function, C_c, which operates on the values supplied by an assignment function. The idea is that the counterpart function may map one individual to another with equal ordinary referential standing. When there are not multiple individuals that share ordinary referential standing, these counterpart functions will have no work to do, and will simply map an individual back to themselves. But when there are multiple individuals with equal ordinary referential standing, counterpart functions may either map an individual back to themselves or to one of their 'counterparts' with equal ordinary referential standing.

On this model, pronouns P have the following semantics, where ϕ represents semantic features (gender, number, class) semantically encoded as presuppositions of the pronoun's use.

$$P_i^{w,t,c,g} = C_c(g(i)) \text{ if } g(i) \text{ is } \phi.$$

So, for example, the semantic value of *you* might be something like

$$\text{you}_i^{w,t,c,g} = C_c(g(i)) \text{ if } g(i) \text{ is the addressee in } c.$$

Consider how such a semantics might apply to utterances of (26).

(26) You're standing by the edge of the lake.

I noted that someone could use (26) to inform Future Jane that Past Jane is standing by the edge of the lake. But one could obviously also use (26) to inform Past Jane that Past Jane is standing by the edge of the lake. On the current semantics, the extension of *you* in both cases would be Past Jane. But the extension may be arrived at in different ways. In the first case, the assignment function might map the numerical index associated with *you* to Future Jane, since she is the addressee. Then the contextually supplied counterpart function would map Future Jane to Past Jane, who would stand as the extension of *you*. In the second case, the assignment function would instead pick out Past Jane, who is the addressee in that case. The contextually supplied counterpart function would then just map Past Jane back to Past Jane.

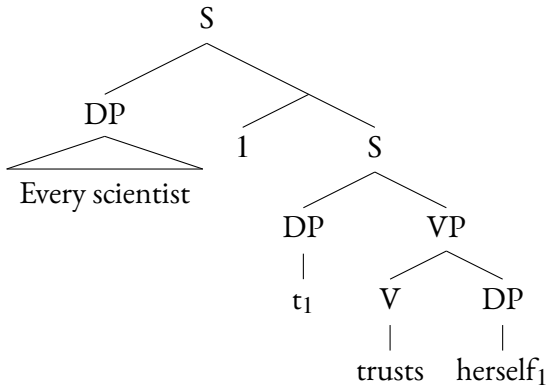
Note the difference between what I've called 'selection' and 'counterpart' functions. A selection function maps a set of objects with equal ordinary referential standing to one of the members of that set. A counterpart function maps one individual to a possibly distinct individual with equal ordinary referential standing. Perhaps we should try to simplify by having only a single kind of function at work in our semantics. But there are obstacles to doing this. We can often find a pair of privileged individuals relevant to the semantics of a deictic pronoun like *you*: for example, the addressee and their contextual counterpart. But it's not obvious we can pick a privileged 'initial' individual to potentially distinguish from their counterpart when we use a proper name. This makes it harder to extend the counterpart machinery to names. Conversely, it will turn out that extending the 'selection' machinery to pronouns may greatly complicate accounts of binding. This will hopefully become clearer shortly. For now, I'll just help myself to both selection and counterpart functions, acknowledging that we may later want to somehow consolidate the semantic machinery.

Let's turn to the case of semantic binding. Consider (28), used in a setting in which two versions of Jane and two versions of Mary encounter each other (assume that all

versions count as scientists). On a familiar (but not uncontroversial) treatment, (28) has a form like (28′), as a result of quantifier raising in logical form.[32]

(28) Every scientist trusts herself.

(28′)



(28) has asymmetric readings on which all four scientists are claimed to trust their counterparts. Whether we can properly account for this will depend on how we treat the semantics of traces. On a standard derivation for (28′), traces behave semantically like pronouns, and at a node dominating a numerical index, we employ a predicate abstraction rule, which binds the numerical index, yielding a property of individuals.[33]

(PA) If α is a branching node with daughters β and γ, where β dominates only a numerical index i. Then $\alpha^{w,t,c,g} = \lambda x_e.\gamma^{w,t,c,g[i \to x]}$.

If we apply these rules on our current proposed semantics for pronouns, then the extension of the embedded S node, at an index-context-assignment triple, is 1 just in case (ignoring constraints on gender and number)

$C_c(g(1))$ trusts $C_c(g(1))$ at t in w.

So the value of the dominating node after predicate abstraction would be

$\lambda x_e.[C_c(x)$ trusts $C_c(x)$ at t in $w]$.

[32] See Heim & Kratzer (1998) Ch.7 for a discussion of this treatment and alternatives.

[33] Cf. Heim & Kratzer (1998) p.186. In the following definition, $g[i \to x]$ is the assignment which differs from g at most in assigning i to x.

This is basically the property an individual (like Past Jane, say) has if her c-relevant counterpart trusts her c-relevant counterpart. This isn't quite what we want though, for two reasons. First, as it stands, this is a symmetric property. Second, to the extent context does any shifting, it would seem to allow that Past Jane could have this property because Future Jane trusts Future Jane. And that doesn't seem to be something we should allow.

The first problem will be avoided by allowing for intrasentential contextual shifts.[34] But the second will not: it seems to arise because we simply extended the semantics of pronouns to that of traces. It is customary to treat these on a par. But in cases of context-shifting interpretations with quantifier raising, we intuitively want the value of the trace to be insensitive to context, and only the value of the pronoun to be flexible. We can achieve this by just dropping the counterpart function from the characterization of the semantic value of a trace. So, where t_i is a trace:

$$t_i{}^{w,t,c,g} = g(i).$$

If we do this, the value of the node dominating the rightmost S will be

$$\lambda x_e.[x \text{ trusts } C_c(x) \text{ at } t \text{ in } w].$$

This is just what is desired. This is the property an individual like Past Jane has just in case Past Jane trusts her c-relevant counterpart, which may be Past Jane or Future Jane depending on the context. So, provided we treat *every* as a binary generalized quantifier, (28′) will have the truth conditions that each of Past Jane, Future Jane, Past Mary, and Future Mary all trust their c-relevant counterparts. This may mean that Past Jane trusts Past Jane, Future Jane trusts Future Jane, etc., or it may mean that Past Jane trusts Future Jane, Future Jane trusts Past Jane, etc., depending on context. It may have further readings if context can make more fine-grained kinds of counterpart mappings, but I won't explore the issue of whether this can occur for (28) here.[35]

---

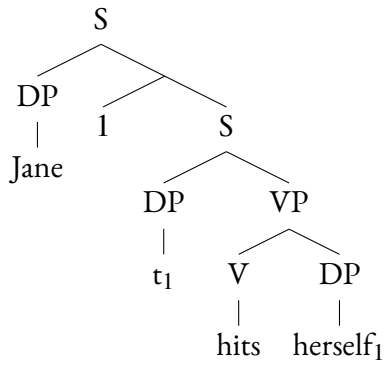
[34] Cf. Braun (1996).

[35] Though see (24) on p.21 for an example where context seems to select counterparts 'differentially'. The unmodified semantics of traces will also help get appropriate readings for relative clauses, provided they too involve predicate abstraction that binds a trace (cf. Heim & Kratzer (1998) Ch.5). Quantified claims will also have readings where context influences the domain of quantification. On the current semantics, this will be accounted for by the ordinary mechanics of quantifier domain restrictions (see Stanley & Szabó (2000) for a discussion) provided, as I've been assuming, the entities of type e include those like Past Jane or Future Jane.

More or less the same machinery can be applied to ordinary cases of syntactic binding like our original (2), assuming that it too contains predicate abstracting operators at the level of logical form as in (2′).[36]

(2) Jane$_1$ hits herself$_1$.

(2′)



The derivation for (2′) yields as an extension

$$\lambda x_e.[x \text{ hits } C_c(x) \text{ at } t \text{ in } w](S_c(\textsc{jane}(\langle w, t\rangle)))$$

This means that (2′) will have value 1 just in case

$$S_c(\textsc{jane}(\langle w, t\rangle)) \text{ hits } C_c(S_c(\textsc{jane}(\langle w, t\rangle))) \text{ at } t \text{ in } w$$

(2) is predicted to have at least four truth-conditionally independent readings, depending on context, as desired: there are two possible ways the selection function may disambiguate *Jane*, and then two further ways the counterpart function may pick out a counterpart of her to be the object of the hitting.

So much for a first form of implementation. I began by reminding that our options for coping with the puzzle were narrowed to rejecting one of (A) or (D), and have so far explored rejecting (D).

(A) The only relevant interpretable constituents of (2) are *Jane*, *hits*, and *herself*.

(D) *herself* in (2) corefers with *Jane*.

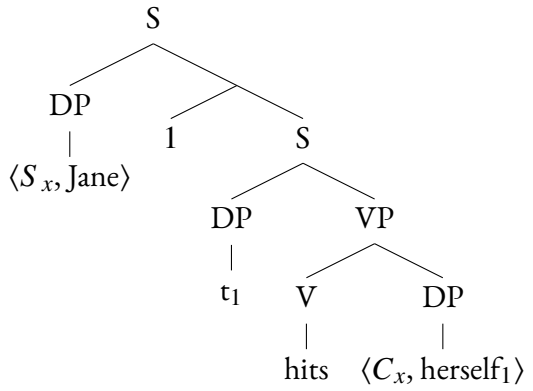
[36] See Heim & Kratzer (1998) Chs. 8–9 for a discussion.

In the foregoing semantics, the referent of *Jane* varies independently of that of *herself*.[37] But we can explore a relatively similar semantics in which we instead deny (A) by enriching the logical form of (2) with variables for selection or counterpart functions.

On what may be the simplest view of this kind, the variables in question cohabit terminal nodes with names and pronouns.[38] So the logical form of (2) might look something like (2″).

(2) Jane hits herself.

(2″)



We begin by giving simple lexical entries for the name and pronoun.

$$\text{Jane}^{w,t,c,g} = \textsc{jane}(\langle w, t\rangle)$$

$$\text{herself}_i^{w,t,c,g} = g(i)$$

Variables that take selection or counterpart functions as values are resolved by context.

$$S_\alpha{}^{w,t,c,g} = S_c$$

$$C_\alpha{}^{w,t,c,g} = C_c$$

[37] Well, strictly speaking *herself* was treated as semantically bound, so that the question of whether it is coextensive, in context, with *Jane* is moot. But the fact that the value of the anaphor wasn't tied to that of the trace left by *Jane* is what accounted for asymmetric readings.

[38] Cf. the treatment of treatment of variables cohabiting terminal nodes with predicates in the treatment of contextually mediated quantifier domain restriction in Stanley & Szabó (2000), which adapts a treatment of contextual variables cohabiting terminal nodes with determiners in Westersthl (1985). I'm open to other placements for the selection/counterpart nodes—for example as sisters to the names or pronouns. The only concern is that enriching structure in those ways may call for special justification. At the very least, any proposals should avoid disrupting argument structure for the reasons given in §3.

We then have a separate rule that determines the value of the ordered pairs at terminal nodes in terms of functional application of the semantic value of the variable to the semantic value of the name or pronoun. This ensures that the value of the ordered pair matches what we earlier had treated as the value of the name or anaphor alone.

$$\langle S_x, \text{Jane}\rangle^{w,t,c,g} = S_c(\textsc{jane}(\langle w, t\rangle))$$

$$\langle C_x, \text{herself}_i\rangle^{w,t,c,g} = C_c(g(i))$$

Thus we'll get the same predictions as beforehand. But, importantly, on this view and unlike on the previous one, the value of *herself* covaries with that of the trace left by *Jane*. That is, we can accommodate coreference for the reflexive, since the semantic burden of shifting the extension has been handed off to the selection/counterpart functions which are now a part of logical form. The ramifications of this shift may seem slight, but may matter for some of the lessons we can draw from the data given by asymmetric reflexives.

Let me note at least one last choice point in implementation. In the foregoing approaches, I associated a name like *Jane* with an individual concept—JANE. But there is at least one approach which can avoid associating *Jane* with such a function, and instead work only with individuals of type e. The view I have in mind is formulated against the backdrop of a four-dimensionalist metaphysics, and makes use of the modest syntactic enrichment approach just given. It then takes *Jane* to denote, instead of an individual concept, a spacetime worm that has two human-like spatiotemporal parts in w at t. Let $\underline{\text{Jane}}$ be the name of such a four-dimensional object. Then the theorist just mentioned may be able to get away with giving *Jane* a very simple semantic value such as the following.

$$\text{Jane}^{w,t,c,g} = \underline{\text{Jane}}$$

This view will then allow something like contextually supplied selection functions to pick out spatiotemporal parts of four-dimensional objects, and counterpart functions to map these spatiotemporal parts to each other. I mention this final option since it seems to be a salient strategy that, unlike the other views just considered, is compatible with Millianism.

7 Consequences and Connections

We've uncovered that, in peculiar cases, sentences containing 'broadly' referring expressions of all kinds, including names and pronouns, both bound and free, exhibit a remarkable form of context sensitivity. This sensitivity seems to drive us to accept one of two conclusions about the semantics of these sentences. It may be that we should acknowledge that all the relevant expressions in these sentences are themselves context-sensitive. If so, we are additionally forced to concede that syntactic binding does not require strict co-reference, but rather reference that respects what I've called equal ordinary referential standing.[39] Or, if not, these sentences compel us to enrich the logical form of these sentences to include a plethora of variables that take selection or counterpart functions as values. Either way, context may function to effectively toggle the extensions of broadly referring expressions among objects with equal ordinary referential standing. Aside from its intrinsic interest, what might such a form of context sensitivity have to teach us?

One set of connections concerns the metaphysics of persistence.[40] How do we humans persist through time? Answers to this question divide theorists roughly into three main camps. According to the endurantist, we persist through time by being wholly present at various times. According to the perdurantist, we persist by being spread out in time—by having temporal parts in both the present and the future, say. According to the exdurantist, 'we'—that is, the ordinary objects of reference and quantification—are not four-dimensional objects but time-slices thereof. We persist by being suitably

[39] As I hinted at in n.2, there has long been evidence that strict coreference does not appropriately capture the semantic notion of 'covaluation' that accompanies syntactic binding. But the most familiar kinds of problems for that notion do not motivate nearly as radical a departure as would the phenomena we've uncovered with (2) (again, see Heim (2009) for a discussion of these issues). There have been some challenges in the ballpark of our conclusion that are more controversial: Jackendoff's example of *Ringo fell on himself* (to mean that Ringo fell on a statue of himself (Jackendoff (1992))), or Nunberg's *Yeats did not enjoy hearing himself read aloud* (Nunberg (1979)). But these examples are possible to treat, and perhaps even best treated, as cases of predicate transfer, which allows us to retain familiar semantic constraints on binding (see Nunberg (1995) §7). We've seen in §2 that such a strategy isn't possible for (2). Also relevant are dream-sentences like *I dreamt that I was Brigitte Bardot and that I kissed myself* (a variant on Lakoff (1970)), or similar counterfactuals. Critically, though, my sentence (2) does not involve intensional operators (even tense operators). This enables us to draw out tighter lessons from the data as having little to do with intensionality, thereby bringing into clearer relief the importance of referential conditions in generating our unusual uncovered form of context-sensitivity.

[40] For a helpful survey, see Haslanger (2003).

related to other time-slices of individuals in the future.

The last two views represent an intramural, and largely semantic, dispute among four-dimensionalists. Cases like (2) have two effects in this narrower debate. First, they may muddy the waters a bit. The question that theorists are trying to answer is what we 'ordinarily' refer to with a name like *Jane*. If we take the last of the views described in §6, it's not obvious that this question will have a clear answer. It's compatible with the data I've reviewed, at least, that *Jane* be lexically associated with a unique spatiotemporally extended object, but that this name cohabits a terminal node with a selection function that picks out something more like a time-slice of *Jane* for us to speak of. If this is the correct semantics for a sentence like *Jane sits*, then who among perdurantists and exdurantists has the right of the issue? After all, the logical form of the sentence has elements that make reference to both spacetime worms and time-slices.

But even though the foregoing results may eventually complicate the question of what continuants are, there is an important sense in which the spirit of exdurantism is favored over that of perdurantism. If we are four-dimensionalists, spacetime worms are not the subjects of predication. Something more like stages of individuals seem to play that role. And such stages are the objects over which we quantify in these cases—quantifying over something like spacetime worms under-predicts the number of readings of quantified sentences.[41]

[41] It's worth bearing in mind, as conceded earlier, that we may refer to and quantify over something more like spacetime worms when occupying a 'timeless' perspective—but this is something that exdurantists may be happy to concede (again, see Sider (1996) pp.448–9, but also Moss (2012) for a discussion of some complexities). It's also worth noting that this evidence favoring exdurantism over perdurantism may be stronger than, or at least substantially supplement, the other most prominent positive case for exdurantism, which is to account for intuitions about counting in cases of coincident objects (Sider (1996), Sider (2001) §5.8, Hawley (2001) §5.5). The problem in leaning on counting intuitions about coincident objects is that those intuitions may be better accounted for by an error theory defended in Moss (2012). According to this error theory, untutored speakers presuppose objects don't coincide and tend to produce false counting claims when coincidence arises. And tutored speakers accommodate the false presuppositions when speaking with layfolk to minimize confusion. This view has the advantage of accounting for the intuition that those who come to accept four-dimensionalism on the basis of puzzles of coincidence ostensibly change their views about coincidence or counting in light of those puzzles. As Moss points out, it is less obvious that the exdurantist (or perdurantist) can account for that change in theoretical view in the same way. In the science fiction cases I've been reviewing, by contrast, we find a kind of 'default' mode of discourse about time travelers that presupposes something like an exdurantist treatment (provided the truth of four-dimensionalism). It's not clear, for example, that we change our views about counting objects when we transition from 'ordinary' discourse about time travelers in science fiction to the more theoretical characterization. If so, the case for exdurantism over perdurantism from time travel discourse may have some of the virtues

This is just one example of the implications that our puzzle may have for debates in metaphysics. I suspect there may be others. But for now I want to set such issues aside in order to highlight a much different implication of the phenomena that we've uncovered, which has nothing to do with recherché cases of science fiction, or recondite issues in the philosophy of time. On the contrary, the phenomenon I want to consider is not only uncontroversially metaphysically possible, but actually occurring and completely pervasive: that of *de re* attitude ascription to confused attitude holders. What I want to note is that the behavior of names and other referring expression in such ascriptions seems to parallel that of names and other referring expressions in our cases of 'overlapping' time travel.

Seeing such phenomena as connected may sound like a stretch. But it is not merely that the semantic behavior of referring expressions is mirrored in these two cases. Rather, on reflection, there are important reasons to think that we have more or less the same phenomenon operating in both cases: context-sensitivity arising to cope with an unusual form of referential indeterminacy owing to a plurality of 'equally good' referential candidates. Before explaining why we should think this, let me review the data from *de re* attitude ascription that creates the first-blush semantic parallel.

The cases of interest to me are sometimes called 'double vision' scenarios where a thinker takes one object to be two. Consider a classic instance:

> Ralph has seen Ortcutt at the beach, and recognizes him there as a pillar of the community. But later, Ralph sees Ortcutt in disguise, standing at the end of the bar. He doesn't recognize him and indeed thinks he is a truly shady character—possibly a spy. At time *t* he thinks that the spy continues to stand near the bar, and that the aforementioned pillar of the community sits comfortably at home by the fire.

Now, does Ralph think Ortcutt is sitting (at *t*) or not? Many have found such questions perplexing, and even somehow a little confused. Quine was among them. Indeed, the question I just asked is (with slight modification) none other than that notoriously pressed by Quine (1956) in his classic discussion of *de re* attitude ascription.

A second important bit of data governing double vision scenarios concerns how ascriptions behave when we piecewise 'subtract' the ways a thinker comes to be acquainted with an object that lead them to posit two objects instead of one.[42] If we tell a

of the case from counting coincident objects, without its concomitant vices.

[42] A point noted by many, including, e.g., Kripke (1979).

slightly different story on which Ralph never sees Ortcutt at the bar, then the claim that Ralph believes that Ortcutt is sitting and not standing (at *t*) seems unproblematically true. Similarly, if Ralph instead never knows of Ortcutt as a pillar of the community once seen at the beach, then it seems false to claim that Ralph believes that Ortcutt is sitting, and true to claim that Ralph believes he is not sitting. That is, the initially troubling question in the case of double vision becomes readily answerable once we cut away either 'half' of Ralph's acquaintance with Ortcutt.

A third bit of data concerns the original case, where Ralph has seen Ortcutt twice over in different circumstances. We can seemingly truthfully say that Ralph thinks Ortcutt is not sitting (say) at *t* if the context is right. For example, if I had told my story as follows, it seems to elicit no confusion, nor generate any inconsistency.[43]

> Ralph has seen Ortcutt at the beach, and recognizes him there as a pillar of the community. But at time *t*, Ralph sees Ortcutt in disguise, at the end of the bar. Ortcutt is in fact sitting. But, because he sees him from an odd angle, Ralph thinks Ortcutt is not sitting, but standing and somehow unnaturally contorted and hunched, which only adds to Ralph's sense that this individual is suspicious. As a result of Ortcutt's disguise, of course, Ralph doesn't recognize him and indeed thinks he is a truly shady character—possibly a spy. Instead, at that time, he thinks the aforementioned pillar of the community sits comfortably at home by the fire.

Mid-dialog, there is no trouble interpreting the narrator as continuing to truthfully report when he says that "Ralph thinks Ortcutt is not sitting, but standing and somehow unnaturally contorted and hunched." In context, this seems like a fruitful way of giving us information about how Ralph thinks of 'the person at the bar'.[44]

---

[43] Cf. the example of 'Shorty' in Schiffer (1979), or the 'three-way' ambiguity in [citation omitted].

[44] The point here is stressed, among others, by Stalnaker (1988). But it's worth flagging that even Quine's set-up of his original case of 'double vision' seemingly involves similar such contextual-resolution. Here is Quine, setting up his case with two of the very *de re* locutions whose intelligibility he was concerned, in some sense, to call into question:

> There is a certain man in a brown hat whom Ralph has glimpsed several times under questionable circumstances on which we need not enter here; suffice it to say that Ralph suspects he is a spy. Also there is a gray-haired man, vaguely known to Ralph as rather a pillar of the community, whom Ralph is not aware of having seen except once at the beach. Now Ralph does not know it, but the men are one and the same. Can we say of this *man* (Bernard J. Ortcutt, to give him a name) that Ralph believes

Contrast these three features of ascriptions of belief about whether Ortcutt sits at *t* to three features of assertions to the effect that our time traveler, Jane, sits at *t*. We saw how it was intuitively hard to settle the question as to whether Jane sits or not, when both *Past Jane* and *Future Jane* are about and the former sits while the latter does not. Or, at least, it was hard to do this in a contextual vacuum. But if just one of *Past Jane* or *Future Jane* were absent at *t*, it would have been easy to interpret *Jane sits* as a straightforward truth or falsehood at that time. And, as we reviewed in §5, even when both characters are present, the right context can again make the claim interpretable as either a truth or a falsehood, intuitively switching which of our two characters is spoken about. All this is to say: there is a striking set of parallels that link our two, seemingly unrelated cases of Ortcutt and Jane. On the theory of attitude ascriptions that I favor, these parallels can and should be thoroughly exploited.

Why? The idea is perhaps best brought out if we help ourselves to a controversial, but familiar and widespread model of Ralph's doxastic state, according to which much of the intentional structure of that state can be supplied by a set of qualitatively characterized possible worlds compatible with his beliefs. Consider, in this theoretical context, what the world is like according to Ralph. If we only concern ourselves for the time being with the qualitative features of his belief worlds, they seem relatively easy to describe. In the world as Ralph sees it, there is an upstanding member of the community, who Ralph sees one day at the beach. And in that world, at time *t*, this individual is seated in his home, by a fire. But in that world there is also a spy hunched over at the end of the bar at that same time *t*. And in the world as Ralph sees it, these individuals are numerically distinct. At least: they comprise at time *t* two heads, four arms, and so on.

When we ascribe beliefs to Ralph about particular persons or objects in the actual world, like Ortcutt, it seems like we are characterizing Ralph's beliefs, somehow, as a function of how things stand with individuals in his 'belief worlds'. But when we say that Ralph thinks Ortcutt sits, which of the individuals in Ralph's belief worlds, as I've characterized them qualitatively, must sit for this to be true?

Before answering this question, we can return to consider what happens when we tell the 'split' stories on which Ralph knows Ortcutt in only 'one way' in each: either by seeing him at the beach or at the bar. In the former story, it seems clear that the man

---

him to be a spy? (Quine (1956) p.179)

in Ralph's belief worlds at home by the fire settles what it is correct to say that Ralph thinks of Ortcutt. And in the latter story, it is the man at the bar. If this is right, these two individuals taken individually satisfy whatever conditions an object *o* must satisfy in Ralph's belief worlds to determine what Ralph believes about Ortcutt.

It is a small (though of course controversial) step from the foregoing claim, to one which establishes a direct semantic parallel between Jane's case and Ortcutt's: the claim that the two individuals in Ralph's belief worlds have what I earlier called 'equal ordinary referential standing.' They individually (in isolation from the other) satisfy the conditions to stand as the extension of an ordinary name like *Ortcutt* in Ralph's belief worlds at *t*.[45]

Why the individuals in those belief worlds satisfy those conditions is just as vexed an issue as settling why each of Future Jane and Past Jane satisfy the conditions to stand as the extension of an ordinary name like *Jane*. As before, there are many possible answers, depending on how we spell out the conditions for an object in Ralph's belief world at time *t* to stand as the extension of *Ortcutt*. Is it that the person in Ralph's belief world must be numerically identical to the individual baptized with *Ortcutt* in the actual world? Is it that the person be an epistemic counterpart to that individual in Ralph's belief worlds under a relation of similarity? Or a counterpart under broadly causal-theoretic counterpart relations? Just as with Jane's case, I won't be concerned to answer any of these questions here. The core of the theory I want to defend is indifferent between them. All that is required of the current proposal is that whatever the relevant referential conditions are, they can be, and are, separately satisfied by the individuals in Ralph's belief worlds, at least in the 'split' cases where Ralph only comes to know Ortcutt in one way.

As soon as we acknowledge this, we will have an explanation of two of the more perplexing features of cases of 'double vision'. First, we have an explanation of why, when we gain additional modes of acquaintance with an object or person, we have problematized *de re* attitude ascription about them in a contextual vacuum, rather than simply a proliferation and straightforward 'compounding' of the pertinent *de re* ascriptions. The reasons for this will be analogous to the explanation in Jane's case: When we have a plurality of objects meeting the conditions to stand as the extension, we have a kind of referential confusion, not reference to a plurality or sum. We have a kind of unresolved-

[45] With the caveat that if we have something like covert counterpart functions in syntax, this claim about 'extensions' may need to be slightly adjusted in the ways flagged in n.28.

ness in what it takes for Jane to sit, not a clear case where a lone individual generates true contradictions by sitting and not sitting at the same time.

Second, this would in turn explain why context can resolve the ambiguity in these cases, so that with appropriate cues we can intuitively speak about what one or the other of the men in Ralph's belief worlds is like. For we have already seen that in Jane's case, context is precisely what allows speakers to get beyond the relevant kinds of referential confusion.

On this understanding of attitude reports, all the complexities of *de re* ascription in cases of double vision arise because of a kind of referential indeterminacy that is alleviated by the mediation of linguistic context. The broad outline of such a treatment of *de re* ascription was, to my knowledge, first put forward by Robert Stalnaker.[46] There is, of course, no time to elaborate this view in any great detail here, or to defend it or contrast it with alternatives.[47] Here, I merely want to mention why drawing the connections between my time travel puzzles and those of *de re* ascription might be very important for defending something like Stalnaker's view, and why those connections may help motivate some very unusual, but potentially illuminating, claims about attitude ascription.

One might wonder, even if the analogy between Jane and Ortcutt's case could be upheld, whether we should expect this analogy to be useful in defending something like the treatment of *de re* ascription just sketched. But actually, a case like Jane's may be precisely what is needed to motivate the most controversial aspects of its semantic implementation. On the view of *de re* ascription just adumbrated, almost all interesting facts about attitude ascription are the result of a semantically mediated influence of context. Appeals to context to solve semantic problems in this way are in danger of seeming *ad hoc*. Why think attitude verbs or, worse, all referring expressions are context-sensitive?

Examining the worlds compatible with Ralph's beliefs, qualitatively characterized, gives us a start. It helps give us a sense for where and why the involvement of context might be needed, since it helps us to see that in cases of double-vision, we proliferate candidates to refer to in ascribing *de re* belief. But then why think that this must lead to referential confusion? Why think context could help? And how would it be implemented? Here our time travel cases seem to give us independent motivation for every

[46] See Stalnaker (1988, 2009).

[47] For a discussion of some of the independent empirical and theoretical motivations for the view, in addition to the above citations, see [citation omitted].

claim the view of *de re* ascription makes: that referential confusion is precisely what would result from the over-generation of referential candidates, that context is capable of resolving the confusion, and that the influence of context in this way would be semantically mediated. Moreover, along with all these claims, we can glean some ideas for how the mediation of context might be implemented compositionally. Of course, we may not use exactly the same compositional tools for both phenomena—perhaps, for example, we will locate the context-sensitivity in *de re* ascription partially in the presence of the attitude verb. The point is that the foundational and theoretical basis for the semantic implementation, and the broad modes of their operation, will be the same. As such, time travel cases may give us independent motivation for the most controversial aspects of a broadly Stalnakerian theory—aspects which also constitute its greatest strength.

In addition to providing us with independent motivations for the controversial semantic resources the theory needs, the comparison between cases like those of Jane and Ortcutt can illuminate some some perplexing phenomena with the help of analogies. Let me mention two briefly, in conclusion.

It was important for my resolution of the puzzle about time travel that context can 'reach in' and change the interpretation of names intrasententially, and even shift the interpretations of syntactically and semantically bound variables. This is no less true of attitude ascriptions. I run into Robert Stalnaker after a talk but do not recognize him. Later I report: "I didn't realize that Stalnaker was Stalnaker." I can even approach the philosopher later and tell him to his face: "Sorry I was so oblivious. I didn't realize that you were none other than yourself." I see all these as having true readings that are expressions of my confused mental state, on which I believed the philosopher who works at MIT wasn't standing before me during the conference. In both cases, ambiguity in candidates to settle my beliefs about Stalnaker are shifted by context, resolving first in favor of a candidate philosopher in my belief worlds at the conference not named *Stalnaker*, and resolving next in favor of a candidate philosopher working away in my belief worlds at MIT. In the second case, this shifting contextual resolution persists in spite of syntactic binding. Examples involving semantic binding also seem possible to to devise.[48]

A second point of comparison concerns the role that linguistic context can play in contextual resolution. In particular, the choice among several 'ordinarily' coreferring

[48] Cf. the discussion of 'bound *de re*' readings in Sharvit (2011), Charlow & Sharvit (2014).

expressions can have effects on contextual resolution in a potentially misleading way. Booker at age 30 changes his name to *Zachary*. At 60, Zachary travels back in time to meet himself at 25. Here are claims that can intuitively be true: Zachary is 60 at the time of the meeting. Booker is 25. Zachary is not 25 at the time of the meeting, nor is Booker 60. Booker stands and does not sit during the meeting, Zachary sits but does not stand. The foregoing claims can easily be easily interpreted as truths, especially with a little more context. But this can mislead. One might conclude that that since *Booker* and *Zachary* tend to be interpreted differently in these ways, they have different compositional semantic values. But this doesn't follow. The change in truth-values among claims about Booker and Zachary are best treated as a result of shifting linguistic context. Which names we choose to refer to an individual form part of the context that may influence how that name is contextually resolved in a case of referential indeterminacy.

A key reason to take this line is that changing one's name doesn't change who one is. Zachary was once called *Booker*. Booker would come to be called *Zachary*. More to the point (adopting the atemporal perspective): Booker is Zachary, Zachary is Booker. When Zachary—or Booker—changed his name he didn't change who he was, but only set up a new way to talk about himself, and perhaps created a norm whereby the new name was the more socially appropriate moniker henceforth. As I say, those conventions are merely social, not semantic. But those social conventions may influence our strategies for contextual resolution of candidates with equal referential standing. When our time traveler, X, encounters himself, I could be meaning to capture facts about one of two candidates when I use the word *Booker*—which? I chose *Booker* rather than *Zachary*, and the former is (or was) conventionally used to talk about X at a certain age, and one of our candidates is of that age. So a natural way to resolve the indeterminacy is to take that name to refer to the younger candidate. Note that the procedure here is extremely unusual, involving the choice of name interacting, contextually, with its own semantic value. Also note that the contextual effects here are defeasible. I can equally describe the resulting situation as one in which (a young) Zachary stands in a room with himself, or one in which (an old) Booker does.

So, too, with attitude ascriptions. Newspaper reporters have witnessed the exploits of a gaudily clad alien superhero and named him *Superman*. Among those reporters is a mild-mannered, bespectacled native of Smallville, dubbed *Clark Kent* by his adoptive parents. This is all just recent history. The two individuals, as it happens, are one and the same person. But Lois isn't aware of the identity. We can accordingly say, truth-

fully, that Lois believes that Superman does not wear glasses, and that she believes that Clark Kent does. How? To avoid confusion let's call the lone individual Lois is thinking about *Kal-El*. In the world as Lois takes it to be there are two candidates to settle what Lois believes about Kal-El: one human, one superhuman. These candidates are numerically distinct in the world as Lois sees it. When Lois is claimed to have beliefs about Kal-El, which of these candidates settles what she believes? On the current proposal, since we know *Superman* is regularly used around Kal-El while he takes on a behavior and external appearance most resembling the superhuman in the worlds compatible with Lois's beliefs, it's reasonable to take that candidate to be at issue. Likewise, *mutatis mutandis* for talk of what she believes about Clark Kent. And on the current proposal these interpretations should be defeasible, as they are. Two persons who know almost everything about Superman and Clark Kent, including public ignorance about him can say things like "Lois is among those confused about who Clark Kent is. When she sees him wearing his tights, Lois thinks Clark Kent is an amazingly strong superhero." And so on.

In one sense Lois believes Clark is a superhero, in another she believes he is not. I may seem to be attributing to Lois contradictory beliefs roughly along the lines of, say, Salmon (1986). But I am not—not any more than I'm endorsing the claim that there are actual true contradictions because in one sense Jane sits at t and in another Jane doesn't. Propositions bear their truth-conditions essentially. As such, I claim that *Jane sits* expresses different propositions in different contexts, with the ambiguity tracing to context-sensitivity in *Jane*. Relatedly, I claim that *Lois believes Clark Kent is a superhero* can likewise express different propositions in different contexts, where the context-sensitivity traces to *Clark Kent*, *believes* or both.

An interesting remaining question is whether something like the foregoing view is compatible with Millianism. The answer seems to be: it may be, provided at least that Millianism is compatible with the data from discourse about time travel. I sketched one such alternative near the end of §6. But, intriguingly, even if Millianism fails we should not abandon substitutivity for names, typically treated as a hallmark of a Millian position. In attitude ascriptions, substitution of actually coreferring names is to substitute expressions with the same semantic value, whichever implementation of the involvement of the relevant context-sensitivity we choose. This is so just as substituting two names for the same time traveler, as occurs in *Zachary sits* and *Booker sits*, doesn't change the semantic value of the sentence as a whole, though it may influence the in-

terpretation of each sentence, and so the proposition it expresses, in context. Whether Millianism can survive in addition to substitutivity is a tough question that will depend on a mix of empirical, metaphysical, and other theoretical commitments. But, importantly, a broad theoretical understanding of what leads to the peculiarities of *de re* ascription, and why, turns out to be independent of those details. What is most important is that a resolution of broadly the form defended in §§5–6 is extended to treat the attitude ascriptions. If all this is right, the seemingly exotic puzzle with which we began turns out not to be so exotic after all. Rather it reveals, in a somewhat cleaner fashion, a strange form of semantic indeterminacy or confusion underlying key problems about attitude ascriptions that have been driving some of the most central work in philosophy of language for over a century.

References

Braun, David. 1996. "Demonstratives and Their Linguistic Meanings." *Noûs*, vol. 30 (2): 145–173. [29]

Carnie, Andrew. 2013. *Syntax: A Generative Introduction (*3rd ed.*)*. Wiley-Blackwell. [2]

Charlow, Simon & Yael Sharvit. 2014. "Bound 'De Refi Pronouns and the LFs of Attitude Reports." *Semantics and Pragmatics*, vol. 7 (3): 1–43. [40]

Fine, Kit. 2007. *Semantic Relationism*. Blackwell. [22]

Gerrold, David. 2003. *The Man Who Folded Himself*. BenBella Books. [3]

Gilmore, Cody. 2007. "Time Travel, Coinciding Objects, and Persistence." In *Oxford Studies in Metaphysics, vol. 3*, Dean Zimmerman, editor, 177–198. Clarendon Press. [16]

Haegeman, Liliane. 1994. *Introduction to Government and Binding Theory (*2rd ed.*)*. Blackwell. [2]

Haslanger, S. 1989. "Endurance and Temporary Intrinsics." *Analysis*, vol. 49 (3): 119–25. [18]

Haslanger, Sally. 2003. "Persistence Through Time." In *The Oxford Handbook of Metaphysics*, Michael J. Loux & Dean W. Zimmerman, editors, 315–354. Oxford University Press. [33]

Hawley, Katherine. 2001. *How Things Persist*. Oxford University Press. [34]

Heim, Irene. 2009. "Forks in the road to Rule-I." In *Proceedings of NELS38*, A. S. M. Abdurrahman & M. Walkow, editors. GLSA Publications, UMass Amherst. [2], [33]

Heim, Irene & Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Blackwell. [28], [29], [30]

Horwich, Paul. 1975. "On Some Alleged Paradoxes of Time Travel." *Journal of Philosophy*, vol. 72 (14): 432–444. [1]

Jackendoff, Ray. 1992. "Mme. Tussaud Meets the Binding Theory." *Natural Language & Linguistic Theory*, vol. 10 (1): 1–31. [33]

Johnston, M. 1987. "Is There a Problem about Persistence?." *Proceedings of The Aristotelian Society*, vol. 61: 107–35. [18]

Kripke, Saul. 1979. "A Puzzle about Belief." In *Meaning and Use*, vol. 3 of *Studies in Linguistics and Philosophy*, 239–283. Springer. [35]

Lakoff, George. 1970. "Linguistics and Natural Logic." *Synthese*, vol. 22 (1-2): 151–271. [33]

Lewis, David. 1976. "The Paradoxes of Time Travel." *American Philosophical Quarterly*, vol. 13 (2): 145–152. [1], [4], [16], [21]

—. 1986. *On the Plurality of Worlds*. Blackwell. [21]

Looper film Wikipedia. 2015. https://en.wikipedia.org/wiki/Looper(film). Retrieved June 1, 2015. [3]

Lowe, E. J. 1988. "The Problems of Intrinsic Change: Rejoinder to Lewis." *Analysis*, vol. 48 (2): 72–7. [18]

Markosian, Ned. 2004. "Two arguments from Siderfis *Four-Dimensionalism*." *Philosophy and Phenomenological Research*, vol. 68 (3): 665–73. [16]

Miller, Kristie. 2006. "Travelling in Time: How to Wholly Exist in Two Places at the Same Time." *Canadian Journal of Philosophy*, vol. 36 (3): 309–334. [16], [18]

Moss, Sarah. 2012. "Four-Dimensionalist Theories of Persistence." *Australasian Journal of Philosophy*, vol. 90 (4): 671–686. [17], [34]

Niffenegger, Audrey. 2003. *The Time Traveler's Wife*. Houghton Mifflin Harcourt. [3]

Nunberg, Geoffrey. 1979. "The Non-Uniqueness of Semantic Solutions: Polysemy." *Linguistics and Philosophy*, vol. 3 (2): 143–184. [7], [33]

—. 1995. "Transfers of Meaning." *Journal of Semantics*, vol. 12: 109–132. [7], [9], [33]

Parsons, Terence. 2000. "Underlying States and Time Travel." In *Speaking of Events*, Achille Varzi, James Higginbotham & Fabio Pianesi, editors. Oxford University Press. [6], [17]

Primer Reviews and Ratings – imdb. 2015. http://www.imdb.com/title/tt0390384/reviews. Retrieved June 1, 2015. [4]

Quine, W.V.O. 1956. "Quantifiers and propositional attitudes." *The Journal of Philosophy*, vol. 53 (5): 177–187. [35], [37]

Recanati, François. 2004. *Literal Meaning*. Cambridge University Press. [7]

—. 2012. "Pragmatic Enrichment and Conversational Implicature." In *Routledge Companion to the Philosophy of Language*, G. Russell & D. Graff Fara, editors, 67–78. [7]

Salmon, Nathan. 1986. *Frege's Puzzle*. Ridgeview. [42]

Schiffer, Stephen. 1979. "Naming and Knowing." In *Contemporary Perspectives in the Philosophy of Language*, A. French Peter, E. Uehling Theodore, Howard Jr & K. Wettstein, editors, 28–41. University of Minnesota Press. [36]

Sennet, Adam. 2015. "Ambiguity." In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta, editor. Spring 2015 edn. [5]

Sharvit, Yael. 2011. "Covaluation and Unexpected BT Effects." *Journal of Semantics*, vol. 28 (1): 55–106. [40]

Sider, Theodore. 1996. "All the World's a Stage." *Australasian Journal of Philosophy*, vol. 74 (3): 433–453. [17], [34]

—. 2001. *Four-Dimensionalism*. Oxford University Press. [4], [15], [16], [18], [21], [34]

Simon, Jonathan. 2005. "Is Time Travel a Problem for the Three-Dimensionalist?" *The Monist*, vol. 88 (3): 353–361. [16]

Stalnaker, Robert. 1988. "Belief Attribution and Context." In *Context and Content*, 150–167. Oxford. [36], [39]

—. 2009. "What is De Re Belief?" In *Essays in Honor of David Kaplan*, J. Almog & P. Leonardi, editors. Oxford University Press. [39]

Stanley, Jason & Zoltán Gendler Szabó. 2000. "On Quantifier Domain Restriction." *Mind and Language*, vol. 15 (2 and 3): 219–261. [29], [31]

Van Inwagen, Peter. 1990. "Four-Dimensional Objects." *Noûs*, vol. 24 (2): 245–55. [18]

Westerstahl, Dag. 1985. "Determiners and Context Sets." In *Generalized Quantifiers in Natural Language*, J. van Benthem & A. ter Meulen, editors, 45–71. Foris, Dordrecht. [31]