

# The Morality of Blackmail

James R. Shaw

*Draft, please don't cite*

Blackmail raises a pair of parallel legal and moral problems, sometimes referred to as the “paradox of blackmail”.<sup>1</sup> Why it is sometimes legal merely to threaten an individual with an action one could legally carry out (e.g., to reveal their marital indiscretions to their spouse), and legal merely to request money from them, but illegal to do them both jointly? Likewise, why is it sometimes morally permissible merely to threaten an individual with an action one could permissibly carry out, and morally permissible merely to request money from them, but morally impermissible to do them jointly?

I propose to address the second, moral problem by bringing instances of blackmail under a general account of wrongful coercion. According to this account, and contrary to the intuitions which give rise to the paradoxes, the blackmailer’s threat to release harmful information is almost never morally impermissible unless it is likewise impermissible to carry out the threat. The important work to be done in defending this account is the delicate matter, in particular cases of blackmail, of identifying the special kind of wrong involved in the release of the relevant information. Carefully doing this also resolves a number of other puzzles about blackmail—for example, why profiting from a threat to reveal sensitive information can sometimes be impermissible, even though accepting identical payment from the same person for the same restraint can be permissible if the payment was independently offered.

My account makes no immediate pronouncement on the corresponding, and much more discussed, legal versions of the puzzle.<sup>2</sup> What it does show is that the core moral case for the illegality of blackmail is precisely the same moral case for the illegality of the would-be threatened behavior that is presently legal. This suggests the presence of a genuine tension in existing blackmail law which we might only be able to overcome by legal revisions.

---

<sup>1</sup>An early discussion of the legal problem, which has been the primary focus of the literature on blackmail, is found in Williams (1954), though without mention of ‘paradox’. As Clark (1994) has pointed out this label might be a bit sensational given that many pairs of actions which separately are legal and permissible can jointly be illegal or impermissible—for example, drinking and driving. The real force of the ‘paradox’ is to bring out a nest of problems in accounting for the illegality or wrongfulness of blackmail, especially those that make it difficult to think of it as a wrong arising from the combination of two permissible actions in the same way as for drinking and driving. It is these which are my primary preoccupation here.

<sup>2</sup>See Lindgren (1984) and Wertheimer (1987) for helpful reviews and criticisms of the wealth of positions in the literature.

# 1 The Paradoxes of Blackmail

## 1.1 Simple Cases: Coercion and Cruelty

I'm going to treat blackmail, definitionally, as a special impermissible type of coercion: it involves the communication to an agent of contingently enforced sanctions in the effort to impermissibly shape that agent's behavior. Blackmail, again definitionally, is a type of coercion distinguished by characteristically involving the threatened disclosure of harmful information. One can use the term "blackmail" more broadly than I do to include some permissible threats of information release, or even other acts of extortion disconnected from the release of information altogether. One could also use the term "coercion" more narrowly than I do so that all types of coercion are impermissible. In the context of this paper, these are mere terminological choices.

Since blackmail is just one special kind of wrongful coercion, a natural way to understand the impermissibility of blackmail involves integrating it into an account of impermissible coercion more generally. Coercive announcements are designed to affect an agent's deliberation by overtly attaching sanctions to the actions available to them: the announcements let the agent know that some actions come with new costs, or costs that weren't antecedently known. When the mugger announces "your money or your life", he tries to convince you that walking away from the interaction with money in hand now brings about your death. When the father tells his child "no dessert unless you finish your meal" he tries to convince her that leaving brussels sprouts on her plate now costs her rights to ice cream later.

Since coercive announcements are designed to affect deliberation, a natural account of wrongful coercion suggests itself, which I'll call the simple account of wrongful coercion.

**Simple Account:** A coercive announcement is wrongful if it attaches a sanction to an option in an agent's deliberation which that agent is entitled to deliberate with free of that sanction.<sup>3</sup>

Some things need to be said straight away about what *Simple Account* is *not* doing. First, I haven't even yet said what counts as a coercive announcement to begin with. What, for example, makes an announcement attach a *restrictive* or *punitive* sanction in deliberation in the way characteristic of coercive announcements (as opposed to those, for example, which attach a 'benefit' or which open up new options in deliberation)? I won't be addressing this question here. Since I'm primarily interested in cases where it is in clear view whether an announcement is coercive or not, I'll let the notion of whether an announced sanction is 'restrictive' operate on an intuitive level.<sup>4</sup>

---

<sup>3</sup>This conditional might be a little too strong: it might be that the announcement is only wrongful if it has an *actual effect* on deliberation—otherwise the agent suffers no 'deliberative harms'. I'll ignore this complication here.

<sup>4</sup>For a recent account of what makes an announcement restrictive in this way, see [citation needed]. Again, some use the term 'coercion' to subsume announcements that add 'benefits' or reveal new options in action. As always, this is primarily a terminological choice.

Second, as regards explanations of wrongful coercion, *Simple Account* overtly passes the buck in what may seem like a theoretically dissatisfying way: it requires an independent account of when an agent is ‘entitled to deliberate with an available action sanction-free’. What are our rights to deliberation, and where do they come from? The simple account doesn’t provide guidance in these matters on its own, and in fact I won’t try to supply any general answers to these questions here.

The reason for this is that the simple account of wrongful coercion combines with a broad principle governing entitlements to deliberation which has obvious intuitive merits, and will suffice to cope with almost all cases I want to treat in the course of this paper.

**Permissible Deliberation:** If one is entitled to deliberate with the possibility of performing action *A* at all, one is entitled to deliberate as if one could perform *A* free of any sanction on its performance that would constitute a prospective wrong.

The idea behind the *Permissible Deliberation* is as simple as it is compelling: One is entitled, definitionally, not to be prospectively wronged. Accordingly, if one has rights to deliberate at all, we would expect one has an equal entitlement to deliberate *as if* one *would not* be wronged. Put in another way, anyone who owes you, on pain of wronging you, not to do *X*, equally owes you the right to deliberate with an otherwise permissible choice as if they will not do *X* following that choice. The question of how we get any rights to deliberation at all is no doubt an important and challenging one, but the intuitive need for such rights, and the intuitive justification for the *Permissible Deliberation*, should I hope be clear enough to permit my sidestepping such questions here.

When combined the *Simple Account* and the *Permissible Deliberation* classify a wide range of coercive announcements as impermissible. Suppose that *B* is entitled to perform *Y* and not-*Y* and so, presumably, is entitled to deliberate as to whether to do *Y*. Then we have the following constraint on coercive announcements:

- (C) If *A*’s doing *X* while *B* does *Y*, or *A*’s not doing *X* while *B* doesn’t do *Y*, is impermissible, then it is impermissible for *A* to communicate to *B* that they will do *X* if and only if *B* does *Y*.

(C) tell us that, for an important range of cases—those in which an agent had antecedent rights to perform any of several actions—coercive announcements attaching sanctions to those actions essentially inherit their impermissibility directly from the impermissibility of any of the sanctions. This is why the mugger’s announcement is impermissible: it is impermissible for the mugger to shoot you if you walk away, cash in hand. Though the above condition only provides a sufficient condition for wrongdoing, if it is the primary source of wrongful coercion, it also gives a partial explanation of why parents are permitted to threaten their children with no dessert: it is arguably permissible for parents to deny their children dessert if they haven’t finished their dinner.

So it is likewise permissible for them to threaten to do so.<sup>5</sup> In this paper, I want to remain agnostic as to whether (C) is the only principle needed to explain all instances of impermissible coercion, but I will assume that it is the primary principle of its kind.

The account of coercion I'm appealing to here faces a number of important challenges, most importantly that the account threatens to explain too little. It is often claimed that principle (C) on its own classifies far too few cases of wrongful coercion. In fact, one of the classic examples used to push this objection is the case of blackmail itself.<sup>6</sup> If the paradox of blackmail starts from correct premises, it is sometimes impermissible to threaten to do something that would otherwise be entirely permissible. For example, it seems easy to dream up cases in which *A* becomes aware of *B*'s marital infidelity in a way that makes it perfectly permissible for *A* to inform *B*'s spouse, though it is impermissible for *A* to use the information as a way to extort money from *B*.

What I'd like to do now is build a case that appearances here are deceiving, and that most intuitively impermissible cases of threatened information release can be accounted for solely with the principle (C). This is meant to form part of a 'mutually reinforcing' case for adopting both the simple account of wrongful coercion and my proposed explanation of the particular wrongs of blackmail together. The way I plan to do this is by bringing into relief a subtle form of moral turpitude which I'll call, for lack of a better term, *cruelty*. I'll then claim that if we closely examine the threats to release harmful information in the allegedly 'paradoxical' cases, we'll see that the genuine cases of blackmail (those which are impermissible), and only those, involve the moral wrong of cruelty in carrying out one's threat.

The kind of moral defect present in these cases that I'm calling "cruelty" doesn't only arise in cases involving coercion, threats, offers, or even special rights to information. Consider Fred, a landlord, who rents his apartment on a renewing monthly lease to Lucy. Fred lets Lucy know ahead of time that he may terminate their arrangement any month, provided he gives at least a week's notice. One month, Fred becomes acutely aware that deciding to disallow Lucy from renewing her lease that particular month alone with only a week's notice will make it extremely hard on Lucy. The difficulty in finding a new apartment in time to move, in addition to causing her great distress, would exacerbate her finances which are temporarily troubled through no fault of her own. Suppose Fred nonetheless decides, on a whim, to terminate Lucy's contract instead of waiting a few weeks. Fred's grounds for the change? He's grown a little tired of having Lucy in the building, and feels it might be nice to see a change of faces.

Fred may be perfectly within his legal rights to do this, but his action is unconscionable. Even though the apartment is Fred's to do with as he pleases, including changing renters as often as he likes, and even though he gave Lucy ample warning, his actions demonstrate an impermissible disregard for Lucy's well-being. His action is *cruel* in the following special sense: it consists in an

---

<sup>5</sup>Depending on the details of the case, the story of why such threats are permissible might be a little more complex. See §2.2 for a discussion of coercive threats aimed at punishment.

<sup>6</sup>See [citation needed].

agent knowingly performing an action that does harm to another which is not offset by an appropriate furthering of other valuable ends (typically valuable to the agent perpetrating the harm). I add that the distress Fred causes Lucy is not ‘offset’ because the cruelty—the impermissibility—of Fred’s action may disappear so long as Fred has appropriate ends suitably furthered by his action. Suppose Fred has a renter who is willing to pay him more than twice what Lucy does, and Fred himself is presently in dire financial straights. Or suppose Fred needs the apartment to move into himself, since his own had to be fumigated on short notice. These are the kinds of considerations which make Fred’s actions *less* cruel, and may ultimately make Fred’s action morally permissible. Importantly, the harm to Lucy in these cases is not lessened. Fred is nonetheless morally permitted to cause this harm without Lucy’s (additional) consent both because the property is his and because he has his own rightful goals to further. A key reason Fred is not actually permitted to terminate Lucy’s contract is because the end of seeing new faces in the apartment building doesn’t excuse bringing serious, and easily avoidable, harms to another person.

Saying precisely how harms and furthered-ends are weighed in assessing an act for cruelty, and what counts as a legitimate *kind* of end to offset a harm, are both delicate matters, so I won’t pursue them here. All that I’ll need for my claims to follow are that cruelty exists as a form of moral impropriety and that one ascertains whether someone has been cruel by attending to two features—the harm knowingly done and the ends furthered by the one causing the harm—and weighing them against each other *somehow*.

A very important fact about cruelty comes from a corresponding fact about value: the value an agent gets by furthering their own own legitimate ends can be sensitive *purely* to their mindset. This means that whether an agent is cruel can be sensitive purely to their mindset as well. Consider: Fred is the last living collector of rare kinds of natural petrological-fungal formations. Because Fred is the only person who can appreciate their beauty and great rarity, these formations have absolutely no value to anyone else. The prize formation Fred has been after for twenty years and that would complete his collection, however, is presently held by someone leaving the country soon, who knows very well that to Fred it is absolutely priceless. The only way Fred can acquire it, and fulfill his lifelong dream, is to evict poor Lucy to get added rent before the formation leaves the country with its present owner. Certainly, with the details of Fred’s passion and Lucy’s plight adjusted in the right way, this might make Fred’s doing so both intelligible, and permissible. Fred might have set the terms of his lease with exactly this kind of thing in mind. But change the scenario *simply* by altering Fred’s mindset and this may no longer be the case. Suppose Fred had continued to collect his formations, but over the years largely lost his interest in doing so. Now he thinks completing his collection is about as worthwhile as finishing the morning crossword puzzle. In this case, evicting Lucy may be unforgivably cruel. In this way, the presence of cruelty is highly sensitive to facts about agents’ own perception of, and receptivity to, the value of the ends they are pursuing.

So much for what I’m calling “cruelty”. My thesis is this: that in typi-

cal ‘paradoxical’ cases, blackmail is impermissible because the release of the relevant information is impermissibly cruel (provided its release is not already impermissible for other more obvious reasons). The cases where the release of the information *seems* permissible tend to be ones where the blackmailer has a default entitlement to release the information (e.g. legally), which would only be trumped by the cruelty the release of the information would involve, much as Fred has a default entitlement to switch tenants in his apartment that can sometimes be morally trumped by the harms that would result.

To begin to argue for this thesis, let’s start with a pair of specially tailored cases.

**Case 1A.** Alva is taking a picture for artistic purposes in a public place and happens, while doing so, to get a sharp image in the background of Bea entering a shop of ill repute. Bea happens to work at a position where public image is important to company profits. The board at Bea’s company already knows a little about pertinent aspects of her private life and deems her activities (appropriately let’s say) irrelevant to her job performance. If, however, the information about Bea’s behavior were to be made general public knowledge, Bea would have to be fired to placate consumers. When Alva is considering which photos to include in a collection of his for an upcoming book, he finds the relevant photo of Bea of little interest and is largely indifferent as to whether to include it or several other photos taken at the same time from which Bea is absent. Alva is about to mark the photo with Bea as one to include in the collection when he recognizes Bea in the shot and realizes that if the photo is printed Bea’s will be fired. Alva eventually shows the photo to Bea, explains the situation, and announces “I will be printing this photo in my upcoming collection, unless I’m paid \$10,000.”

Intuitions about this case and the others I will present often depend on details which I haven’t spelled out—details about Bea’s relationship to Alva, and details about how Bea presents herself in public, to take two examples. But on reasonable ways of filling in those details I take this to be a case where many people, myself included, would be inclined to say that what Alva does is impermissible. Trying to use the photo for profit in this case constitutes an objectionable attempt by Alva to manipulate Bea for personal gain. This case is also very much like those that are said to give rise to the puzzles about blackmail I mean to address. The picture and its reproduction, as a matter of property, clearly belong to Alva. The image was made in a public place. Isn’t it clearly permissible, then, for Alva to print his photo in his collection? Before answering this question I want to consider another case.

**Case 1B** As in case 1A, except Alva has a very different reaction when looking at the photo taken. “This” Alva says to himself “is a masterpiece.” Alva sees in his photograph a composition whose perfection is the product of a unique interaction between chance

and special, idiosyncratic choices he brings to his photography. This may, Alva thinks, be one of his two or three favorite photos he has ever taken—the kind that he had always hoped to produce as a photographer and the kind he would hold out as a representative of what he hopes to achieve with his work. Enthusiastically, Alva is about to mark the photo as one to include in the collection when, as before, he recognizes Bea in the shot. Alva hesitates, thinking of the harm that would come to Bea as a result. Leaving out the picture would be foregoing an opportunity that comes but a few times in one’s life. Alva is ultimately willing to make the sacrifice on Bea’s account, but is hoping for *some* compensation. Alva eventually shows the photo to Bea, explains the situation, and announces “I will be printing this photo in my upcoming collection, unless I’m paid \$10,000”.

On reasonable ways of filling in unmentioned details, I take this to be a case where many people, myself included, would be inclined to say that what Alva does is *permissible*. Indeed, Alva’s behavior on slight elaboration may plausibly be considered self-sacrificial and *supererogatory* if he considers the \$10,000 small compensation for his artistic losses. Consider, for example, that it is perfectly compatible with what I’ve said, and perhaps even likely given it, that Alva sincerely hopes that Bea decides not to pay so that he may publish his photo with peace of mind having given Bea a fair chance to compensate him. If this is right, a question arises: Alva has made the same contingency announcement, to the same individual, to the release the same information, with equally harmful effects. So why might that contingency announcement be impermissible in 1A but permissible, and possibly even supererogatory, in 1B?

The only difference between the cases, by design, is Alva’s relationship to the photograph and its eventual release. A tempting way to state how this affects permissibility is by saying that in case 1A a *threat* is being made while in case 1B we have an *offer*. I think this is, in a sense, correct but ultimately uninformative. After all, what *makes* the former a threat and the latter an offer? And why is the offer permissible and the threat not? One of the simplest ways to explain what constitutes a threat is to appeal facts about a ‘baseline’ of well-being. An announced action is a threat if it would bring one’s addressee’s well-being below the baseline, and it is an offer if it raises them above the baseline. The simplest way of specifying the baseline is as the level of well-being that would have obtained had the threat/offer not been made or carried out. But of course, I’ve stipulated the details of the case so that this won’t work. The baseline on that specification is identical for Bea in each case: quite bad, as the photo would have been published in both cases. Other ways of specifying the relevant baseline are controversial.<sup>7</sup> And even once we have specified that baseline we’ll still be faced with the challenge of saying how this lines up with questions of permissibility and impermissibility. After all, offers *can* be impermissible, and threats permissible.

---

<sup>7</sup>See [citation needed].

So let's set aside the strategy of appealing to the language of threats and offers for now. After all, the point of raising case 1B to contrast with 1A should be relatively clear: this pair of opposing cases bears obvious similarities to my earlier considered cases of Fred *qua* landlord and collector. In all cases, judgments of the permissibility and impermissibility of certain actions are clearly tracking facts about how an agent relates to the value of various ends they can pursue. But I want to claim more: the grounds for the impermissibility of Alva's action in 1A are essentially the *same* grounds for the impermissibility of Fred's original treatment of Lucy. Showing why will give us a better picture of what's going wrong in 1A and of why nothing is going wrong in 1B. And the way to tie all these cases together is precisely the account of wrongful coercion given in (C).

To see this, let's focus on case 1A again. I ended discussion of that case with a question. It was permissible for Alva to take the picture he did, and he has rightful ownership over it and its reproduction. So isn't it permissible for Alva to print his photo in his collection? There are several senses in which it is fair to say this. Alva certainly has a legal right to do so. It would even be fair to say that Alva has a 'default entitlement' to do so—I'll say a little more about this in §1.3. But in case 1A it is impermissible for Alva to publish the photograph *once Alva has seen the harm it will do to Bea*.

To see this, simply set aside the issue of threats altogether. Imagine, for example, that it never occurs to Alva to use the photograph to extort money from Bea. Recall that, in 1A, Alva was basically indifferent as to whether or not to include the photo in his upcoming collection. It was a stipulation of the case that Alva needed only the slightest reason to include one of several different photos in the collection instead. Now we imagine in these circumstances that Alva stumbles on the fact that the photo would destroy Bea's career. Alva has no grudge against Bea, no reason to think that Bea deserves to lose her job, or that the information that leads to that consequence ought to be revealed. Alva sees nothing *good* that comes of releasing the information. If we keep these facts in view, and Alva nonetheless does not make the minimal effort to substitute a separate photo, I think it becomes relatively clear that Alva has done something impermissible. The action here is just an instance of what I earlier called "cruelty": Alva performs an action that knowingly causes great harm to Bea without being offset by the value of the ends that Alva furthers through his action. The harm, in combination with Alva's relative indifference, 'trumps' any rights Alva has to do what he'd like with his picture in this instance.

I noted earlier that assessing an action for cruelty inherently involves attending to facts about an agent's state of mind, since those facts can shape what kinds of value an agent can get from pursuing various courses of action. This is the value that we weigh against the harms they cause in testing if an action is impermissibly cruel. Switching from case 1A to 1B simply seems to give another case where the 'offsetting' value of an agent's action changes because of a different attitude to that action or a different receptivity to the good the action may supply. In this case, Alva attaches a great value to distributing the photograph which happens to picture Bea, since the photo is now seen as something

of independent value and distributing it part of a worthwhile lifelong goal of creating and sharing objects with that value. As such, the cruelty in releasing the information vanishes, offset by the newly gained value the distribution of the photo has for Alva. Since the photograph is Alva's property, physically and intellectually, this makes Alva's *independent* grounds for releasing the photo sufficient for the permissibility of that course of action.

What does this teach us? If my diagnoses of these cases are correct, we seem to find that *intuitions about the permissibility of a threat are directly tracking facts about the cruelty of the action threatened*.<sup>8</sup> This means that the intuitions about the permissibility of a threat in these cases are, after all, tracking the intuitions about the permissibility of carrying out that threat. It is easy to overlook this because cruelty is, as I've already noted, a fairly subtle kind of wrong: it can arise in cases where one has a default discretionary authority to pursue certain actions, and what makes it arise can involve fairly subtle facts about the mentality and values of the agents performing those actions. I'll have more to say about how subtle facts about cruelty play a role in generating the apparent paradoxes of blackmail in §1.3. For now what's important is to see how we can explain the wrong involved in 1A—why it constitutes a case of blackmail—by appeal to no more than (C).

Though cases 1A and 1B provide just one example, many of the most common cases alleged to exhibit the 'paradoxical' character of blackmail can be treated analogously. When it seems genuinely permissible for someone to release harmful information, we need to ask what they would gain by doing so, and whether this gain offsets the harm caused by the release of the information. When we do this, I claim, we find again and again that our judgments about the permissibility of threats tracks our judgments about the permissibility of carrying the threat out. As such, 1A and 1B will serve well at least as starting cases to get the elements of my view on the table.

There are of course some more complex cases of blackmail, which introduce additional complicating moral factors. I'll consider several such cases in §2. Even though the cases are more complex, they are unfortunately not always marginal. I'll eventually argue that even the seemingly simplest and most common case of blackmail involving marital infidelity will turn out to involve interactions between my core theory of blackmail and several distinct moral factors. But before I get to these cases, I want to briefly do two things: to show how the very basic structure of my view so far already gives us the tools to avoid another key challenge facing any account of blackmail, and then to show how my view gives an account of why there has *seemed* to be special problems in accounting for the morality of blackmail.

---

<sup>8</sup>Though I do myself have these intuitions, my case in this paper does not require endorsing them. The only claim is that two kinds of judgments are *equally* good, and tend to come together: judgments about the cruelty of releasing information, and of the impermissibility of using that information as part of a threat or offer.

## 1.2 The Second Paradox of Blackmail

DeLong (1993) has drawn attention to a second puzzle about of blackmail: sometimes it is illegal to request goods as compensation for withholding information, but it is perfectly legal to accept an offer of the exact same goods to withhold the very same information. Why is this so? The parallel moral question also seems to arise: why is there sometimes a moral difference between acquiring goods after a threat, and accepting them after an offer, when the goods are being transferred for essentially the same reasons? Doesn't it substantially change the case if, in 1A, Bea independently searches out Alva, and offers him \$10,000 not to publish the embarrassing photo?

The account of §1.1 gives us some tools for answering the moral version of the second puzzle. The bulk of the work is again done by the simple account of wrongful coercion. Recall that according to this account, impermissible coercive announcements are impermissible because of the inappropriate kind of influence they have on deliberation. In making an impermissibly coercive announcement, one attaches a sanction to an option in an agent's deliberations that they were entitled to reason with free of that sanction. But if this is right, it enables us to see that the particular kind of wrong involved in blackmail simply cannot be present without an *announced* sanction.

This might be easier to see if we first contrast a case where a wrong like that involved in impermissible coercion is made more tangible. Cid is traveling on foot to a neighboring town, and is deciding which road to take: the longer more scenic path, or the shorter quicker path. Dan, out of selfish or malicious motives, tries to influence Cid's deliberation to make it less likely that Cid chooses the shorter path. There are many ways Dan could do this. Perhaps Dan creates an actual obstacle and places it visibly at the start of the shorter path. Perhaps he only creates the illusion that this is so. If Cid is influenced by any of this when deciding what to do, then barring special circumstances Dan will have wronged Cid, probably in a number of ways. How?

The wrong I'm interested in here isn't simply the wrong that Dan does by forcing Cid to traverse the obstacle. Cid is wronged even if (in fact especially if) he doesn't confront the obstacle because he chooses the longer path. That wrong isn't just the added costs to Cid's journey owing to the presence of the real, or illusory, obstacle. If Cid takes the longer path, which turns out to be the better one for unforeseen reasons, Dan's actions were still impermissibly manipulative. Similarly, the wrong is still present if Cid decides to go ahead on the shorter path only to find he needn't traverse any obstacle at all, because it was illusory. In this latter case, the wrong isn't just the wrong of deception: the same wrong is present whether or not the obstacle really is there. The problem is an undue influence of Dan's actions on Cid's deliberations. Moreover if Dan places an obstacle where Cid *can't see it*, and it doesn't ever affect his choices, then the special kind of wrong present in all the cases I've just given disappears.

The problem is this: Cid has a right to pass on both paths unhindered by Dan and, accordingly, seems to have an equal right to choose as if he will not be so hindered. The simple account of coercion says that all wrongful coercion

infringes on essentially this right. It influences deliberation in creating the impression that obstacles attach to acts which they should be free of, whether or not such obstacles are real. This wrong can be perpetrated whether or not the person influenced knows who is influencing them, whether or not they know the influence is inappropriate, and even whether or not they know *that* they are being influenced at all.

The reason I bring up this simple case is to stress the following point: there is no way for Dan to recreate this special kind of wrong to Cid without some action that interferes with, or at least has the potential to interfere with, Cid's thought process. If Dan refrains from any such interference, Cid may end up choosing the path that everyone, including Cid, acknowledges is manifestly worse for him. That wouldn't *necessarily* place any blame on Dan—and if it did, it would be a very different kind of blame than in the cases where Dan directly meddles with Cid's choices. In the case of impermissible coercion, it's the *announcement* (at least prototypically) which constitutes the way of tampering with another's deliberation. So without the announcement, we should expect that the special kinds of wrongs associated with coercion will disappear.<sup>9</sup>

I want to be careful to stress that it is the wrong of impermissible *coercion* that disappears, because other wrongs of a different kind may surface. For example, when an offer is being made to withhold information, it may be made to someone who misapprehends their situation. This may place the offerer under an obligation to rectify that misapprehension. We need to pin down these potential kinds of obligations if we want to properly understand the variant of 1A where Bea independently offers Alva money to not publish the photo. In such a case, is it really permissible for Alva to accept the money? I think our answer here depends on further details. Why is it that Bea is bothering to make this offer? What does Alva think about what Bea is thinking? What is Alva's relationship to Bea?

To see why these questions might matter, let's revisit Cid and Dan's more concrete case. Suppose Cid is deliberating on his choice of paths and is about to choose the shorter one until he sees Dan, who is on better behavior than before, observing him. Cid eyes Dan suspiciously and then starts off on the longer path. Suppose this gives Dan good reason to think that Cid is making this choice only because he believes Dan has tried to place obstacles in his way on the shorter path. I think it's not unreasonable to hold that Dan has an obligation in this case to communicate to Cid that this is not actually so. If not, I think some things might make it more likely that Dan acquires obligations of this sort—for example if he's done many things in the past to foster Cid's mistaken belief. If in such a case Dan allows Cid to go through with his choice, he hasn't impermissibly interfered with Cid's deliberation, but he may have perpetrated a similar wrong: allowing Cid to labor under the misapprehension that his options were impermissibly constrained.

---

<sup>9</sup>Of course there are many different ways of communicating one's intentions that don't involve language use. By placing emphasis on an 'announcement', I don't mean to accord any special status to speech. In fact, the case of Cid and Dan is precisely one in which the wrong of coercion (albeit along with other wrongs) may arise without any speech.

Similar things can be said about the case of Alva and Bea. Suppose Bea believes that Alva plans to publish the photo expressly, and maliciously, for the purpose of defaming her. If Alva knows this, has no such intention, and indeed has no special reason to publish the photo at all, he might be under an obligation to reveal this to Bea before accepting any money. This seems especially likely if Alva has done anything to foster Bea's misapprehension.

Bea could have other kinds of mistaken belief as well. Perhaps Bea mistakenly believes that Alva is in a situation more like that in 1B, and has a great deal to gain by releasing the photo so that she is merely aiming to compensate Alva. This would be akin to a case where Cid thinks (mistakenly) that Dan has some legitimate independent reason for creating an obstruction on the shorter path. Again, on some elaborations of this scenario, it's not unthinkable that Dan acquires an obligation to correct Cid's misapprehension before he chooses. Similarly, Alva may be under an obligation to divulge that he doesn't stand to gain from publishing the photo, and would refrain for less, before any money changes hands. I suspect these last particular kinds of obligations are ones that can be very easily defeated. They involve the kind of information that it seems permissible to willingly hide in bargaining transactions. If we see Alva's acquisition of money in this light, then Alva may be free to demand more money (given the offer) than it is actually worth to him, by fostering the illusion that the photo is of great independent value to him. The topic of what kinds of information we owe to one another in bargaining transaction is another vexed topic that I'll have to leave aside for now.

I have no special commitments as to when the kinds of obligations I've just been discussing actually arise or are defeated. I only want to note that that the presence of some such wrongs is compatible with my primary point: on the account I've given about the impermissibility of blackmail, we can make sense of why it is that a sense of grievous wrong disappears when goods that would have been received after a threat are instead exchanged after an independent offer. In such cases there is no instance of wrongful coercion. This leaves open that there may sometimes be subtler, and generally less severe, wrongs associated with failures to properly disclose relevant information. And this seems true to the cases question.

A final comment on a third puzzle about blackmail, though one which hasn't been distinguished with the label "paradox": why is blackmail wrong even if sometimes people *prefer* to be blackmailed.<sup>10</sup> In our original case with Alva and Bea, Bea may sometimes wish that Alva comes to her asking for compensation when he does not. How is Bea harmed by an outcome that she longs for? Why is there a moral proscription against her wish being fulfilled?

The account of §1.1 shows us what's confusing about this question. In normal cases, like that of Alva and Bea, one of two situations holds. Either informing isn't cruel, in which case Bea actually isn't being harmed, but helped by the request for compensation. Accordingly, there is no mystery about why Bea can

---

<sup>10</sup>The point is often stressed by libertarians who favor legalizing blackmail. See, e.g., Block et al. (2000) p.595.

wish Alva gives her the opportunity to compensate him—this is a permissible, and sometimes generous, action on Alva’s part. Otherwise, releasing the information is cruel. In this case, what Bea doubtless wishes for is that Alva behave a particular way *given* that he is already cruel. It’s no surprise that one could wish for something that would be a wrong to oneself in this kind of way. You might, for example, wish of the serial killer that he merely tortures, but does not kill you. Of course one makes this wish only while ceasing to hope that the serial killer mends his ways entirely. Similarly if Bea wishes Alva to come to her for compensation, knowing he has no independently justifiable reason to release the information, it is because she is worried that he will wrongfully release the information nonetheless. Given this, she wishes that he bring about the lesser harm of extorting money from her. Of course if she let her hopes range more freely, she’d just wish for him to keep the information secret at no cost.

In many cases you may not know whether you are dealing with someone who is cruel or not, but simply worry about the harmful information getting out and want some confirmation that it doesn’t. But there’s no reason to think that if you hope you are given the opportunity pay for this confirmation—whatever the dispositions of the person in control of your fate—you aren’t merely hoping for less than what you are already owed.

### 1.3 The Illusory Case Against the Simple Account

I’ve now given some basic resources to cope with several puzzles about blackmail. I’ve given reasons to think it is actually much more challenging than it is sometimes claimed to come up with a case where a harmful release of information is permissible, but a threat to release that information failing compensation is impermissible. I’ve also explained why in some cases accepting independently offered goods to withhold information can be permissible (or at least substantially less blameworthy) while requesting those goods to withhold the information is not. But like many paradoxes, these puzzles about blackmail need a two-stage resolution. A proper account of any paradox which traces its source to a poor form of reasoning owes not only an account which exposes the poor reasoning, but also explains *why* the original reasoning might have seemed compelling. A key virtue of my position is that it can do this as well.

Often cases used to make vivid the paradoxes of blackmail are somewhat under-described. It’s not atypical to have an example of the first paradox put in even *less* detail than this:

**Case 1C.** Alva is taking a picture for artistic purposes in a public place and happens, while doing so, to get a sharp image in the background of Bea entering a shop of ill repute. Bea happens to work at a position where public image is important to company profits. If the information about Bea’s behavior were to be made general public knowledge, Bea would be fired to placate consumers. Alva shows the photo to Bea, explains the situation, and announces “I will be printing this photo in my upcoming collection, unless I’m

paid \$10,000.”

Is what Alva did wrong? If, instead of extorting money from Bea, Alva had published the photo, would Alva have done something impermissible? There is a temptation to think that the above description gives us all the information we need to answer these questions. And it’s no wonder that this is the case. Even if I’m right that facts about Alva’s attitudes towards the photo can make a difference to our answers to these questions, most ‘work’ done in the example owes to what I have earlier been calling, a bit loosely, ‘default entitlements’. Alva acquired the photo as legal and intellectual property by rightful means. In almost every case, this is a *sine qua non* for Alva’s ability to release the information ethically. Consequently the case is giving us the ‘most important’ information—the kind of information could often be enough to settle the case.

Unfortunately since it doesn’t give us all the relevant information, there is a tendency to ‘fill in’ the extra details of the case. Moreover, it is not only possible, but *reasonable* to fill in those details in different ways when one is asked each of our key pair of questions about 1C. If we ask: “were Alva to ask Bea for money, would he have perpetrated a wrong?” we must evaluate a counterfactual conditional which calls on us to look for reasonable ways of making the antecedent true. We must then ask: what does case 1C have to be like—what is the most reasonable way of filling in the details—to cohere with the fact that Alva asks Bea for the money? Elaborations like 1B seem like contrived, special circumstances, to give as answers to this question. It’s a bit much to simply assume that Alva has some special independent interest in publishing the photo and *that* is the reason he comes to Bea: for compensation. But instead we ask: “If Alva didn’t extort money from Bea and instead published the photo, would Alva have done something impermissible?”, the counterfactual conditional requires us to adjust details in a potentially different way. What would the case have to be like for Alva not to have even tried to extort money from Bea *and* nonetheless to publish the photo? We could imagine the story roughly as in 1A, except with Alva immediately publishing the photo to Bea’s great harm, and at little personal gain, without having made a threat. But it’s important to note that the character we make Alva into is rather peculiar: he’s willing to perpetrate a moral wrong for no personal gain, and is foregoing an equally impermissible action which would profit him quite a bit. It’s much more natural to assume that Alva publishes the photo for independent reasons. Once we make that step, though, our intuitions about permissibility are beholden to a weighing process—one which may require little independent gain for Alva to offset the harm to Bea in order for the action to become permissible.

It’s very easy to conclude from a case like 1C, that there can be circumstances where there can be an impermissible threat to do something otherwise permissible. Getting this intuition, I claim, involves a subtle and reasonable shift when answering two questions, in what might seem like irrelevant facts about Alva’s motives for performing those actions. It’s only once we carefully dissect these cases that we can see facts about such motives turn out, perhaps surprisingly, to be what holds sway over our intuitions whether or not Alva is

undertaking an action, or making a threat to undertake it, permissibly. This is what makes the paradoxes of blackmail seem so forceful.

## 2 Harder Cases

Though I believe the morality of many philosophically tricky cases of blackmail can be illuminated using the distinctions drawn in §1, other cases bring in additional moral complications which can interact in complex ways with the issues of property, entitlement, and cruelty. In this section I want to consider three groups of cases that require a little bit more care than those I've looked at so far:

- (i) threats to release information that one is obligated to disclose,
- (ii) threats to release information as a form of punishment, and
- (iii) threats that constitute permissible 'bluffs'.

### 2.1 Threats to do One's Duty and Cheating Spouses

Until now I've focused on cases where, if we bracketed the issue of the cruelty, threats were issued concerning information that it was *merely* permissible to release (i.e., it was equally permissible to withhold it). This leaves out two important extreme cases: cases where it is obligatory to release the information, and cases where it is straightforwardly impermissible to release the information. The latter are the easiest to account for. If I (say) steal information that belongs to you and threaten to release it to your great detriment, the simple account of coercion captures perfectly well why this is impermissible. But what of the former cases where it is impermissible to withhold the information? The simple account of coercion might seem to predict that it is never impermissible to threaten to release information that one has antecedent obligation to release—and to many this seems incorrect. Consider:

**Case 2A.** In investigating the crime scene of a theft Earl, working for the police, comes across a finger print which he identifies as that of Finn. Instead of bringing these facts to light, Earl confronts Finn with the information, and demands that Finn pay him a large sum of money in exchange for staying quiet.

Many people have the intuition that there are several wrongs perpetrated here, and that *one* of the wrongs is, or should be closely related to, the wrong involved in standard cases of impermissible coercion. Can the account of blackmail I've developed answer such intuitions?

I think so, though again it requires some careful work. After all, if there is a wrong like that involved in coercion here, there are clearly multiple wrongs we need to track. First, there is a wrong connected with the Earl's claim that, given compensation, he won't release information that he should. We see this wrong brought out more clearly in the following variant.

**Case 2B.** In investigating the crime scene of a theft Earl, working for the police, comes across a finger print which he identifies as that of Finn—a close friend of his. After some hesitation, Earl decides to cover up the information for Finn’s sake. Unfortunately the fingerprint is already recorded in the police’s secure database. Earl knows how to erase that information in a traceless way, but it requires some expensive electronic equipment. Earl confronts Finn with the information, and asks Finn to shoulder the cost of the equipment so he can clear him of suspicion.

On reasonable ways of filling in the details of 2B, what Earl does is impermissible, but there is no sense that the impermissibility involves wrongful coercion. What Finn seems to be doing is *expanding* rather than limiting Finn’s options. Finn’s actions seem problematic because, and only because, it is wrong to countenance withholding the relevant information that ought to be released. What’s important is that this wrong is no wrong *to* Earl (except on very special theories of wrongdoing).

2B reveals that if we want to account for the intuitions in 2A, we need to separate out a wrong that is disconnected from harms to Earl, from a potential wrong which is something more like a harm to Earl. Accordingly, the case draws attention to a tempting move on behalf of a defender of my account of blackmail that I think should be avoided. That move is simply to say that all intuitions about wrongdoing in 2A are tracking those supplied by (C) in the following manner: whenever anyone announces that they will do something obligatory if and only if they don’t receive something in return, they are thereby announcing that they will do something impermissible if they do receive something. So the announcement can still inherit its impermissibility from the impermissibility of retaining the information.

The problem with this explanation is not that it fails to apply—the reasoning is fine, and it may explain both why the threat in 2A and the offer in 2B are impermissible. Intuitively, though, there is something problematic about Earl’s relationship with Finn in 2A which isn’t manifested in 2B, and it is this that requires saying a little more.<sup>11</sup>

We can, as always, get clearer on the issue by simply removing the distracting issue of threats and offers altogether. Consider the contrast between 3A and 3B.

**Case 3A.** Earl is a crooked cop, utterly disdainful of the law, with ties to the crime family of which Finn is a member, and comes across the latter’s prints at a crime scene. Earl has always worked to spare members of the crime family from incrimination where possible,

---

<sup>11</sup>The temptation to say that 2A is exhaustively explained as a case involving an impermissible offer is heightened if one focuses on the legal, rather than the moral, dimensions of blackmail. The illegality of Earl’s actions in 2A plausibly depend only on antecedent legal obligations to perform his job. See for example the discussion of “misprison” in Feinberg (1988) pp. 241–245. The potential problems with limiting one’s treatment of the case in this way is appreciated by Wertheimer (1987) p.91.

sometimes hoping to get favors from the family in return. Earl has a great opportunity to do so here. But Earl has never liked Finn, so he expressly divulges the information to his superior, taking delight in Finn's eventual arrest and incarceration.

**Case 3B.** Earl is an upstanding man of the law, finds Finn's prints at a crime scene, and dutifully reports the information to his superior.

Whether Earl is a crooked cop or not doesn't affect what Earl *ought* to do in 3A and 3B. In both cases he ought to bring the information to his superiors. But there's something insidious about Earl's actions in 3A. They don't seem particularly praiseworthy, and they also seem like a kind of affront to Finn. Finn seems to have loose grounds for resenting Earl in 3A that he lacks in 3B.

It's easy to see why this is the case. In 3A, though Earl has good reasons to report the information to his superiors, he doesn't *treat* them as good reasons, and doesn't *act* on them. Instead, he treats the fact that his action will make Finn suffer as his reason to report the information. This (arguably) isn't a reason for Earl to report the information. That's the sense in which Finn has loose grounds for resentment: Earl's actions show a disregard for Finn *given* Earl's antecedent stance towards issues of legal justice.

Why do I bring this up? Ordinary intuitions track some sense in which the relationship between Earl and Finn is problematic in 3A. Whatever account we give of those intuitions will likewise apply to 2A. After all, given (C), precisely the same kind of phenomenon is arising. In 2A, Earl is not giving sufficient weight to the moral and legal reasons for divulging the information about Finn. If he were to release the information about Finn, it would be for the wrong reasons—effectively just to punish Finn. But Earl has no special right to punish Finn *independently* of the wrong Finn committed. The kind of announcement that Earl makes signals this kind of disregard for Finn's well-being. This, I submit, is the 'wrong' Finn does to Earl that is sensed in that case.

There are at least two possible ways of explaining the distorted relationship between Earl and Finn in both 2A and 3A. One is to develop a kind of 'non-ideal' theory of rights and wrongs, where we introduce relativized notions of permissibility and obligation—relativized in the sense that they track what one (morally) ought to do bracketing certain, perhaps very strong, moral reasons. There's some suspicion that we need something like this anyway to account for ordinary language use of conditional "oughts". It's not unusual for people to say things like "What you ought to do is stand alongside Jane and actively help. Given that you're not, you ought at least to steer clear of her altogether." Alternatively we can introduce a division of labor between impermissibility and blame as done by Scanlon (2008), saying that 3A presents us with a case where Earl's actions are permissible (since in the sphere of legitimate action) but blameworthy (because done for objectionable reasons). In the case of 2A, on this view, Earl's action are just as impermissible as in 3B, but additionally blameworthy because of their grounds.

There might be other ways still of filling in the details. The important point is that any tools we use to account for intuitions about 3A will extend to give

as satisfying an account of the ‘added wrongs’ in 2A, and in threats to release information that ought to be released in general. The end result will always be that coercive threats to do what one is obligated to do are both impermissible and show a morally objectionable disregard for the person threatened. It’s just that the moral disregard is not, in this particular case, the *ground* of the threat’s impermissibility. This seems to give precisely the correct diagnosis of such cases, especially when we contrast alternatives, like 2B and 3A, where each of the two individual ‘wrongs’ is factored out independently.

Now that we’ve gained an understanding of threats to do one’s duty, we’re in a position to treat a common case of blackmail whose discussion I’ve deliberately postponed—that of the cheating spouse—though it requires a slight detour. It almost always seems wrong to ask for compensation for nor revealing information about marital infidelity. Why? The answer turns on a number of details. There are clear examples where one may have a duty to inform someone’s spouse of marital infidelity. This is especially true when one has a close connection to the wronged party. There may be admittedly rarer cases, depending on prevailing cultural norms, where one may have a duty not to interfere in strangers’ business. Demanding compensation for retaining information about infidelities will be impermissible in both cases, but receive different explanations that we have just now discussed.

Finally there is a third class of cases where it may be discretionary whether to interfere and inform a spouse about infidelity. But in these cases, almost always, *the only genuine value in informing is connected with providing assistance to the wronged spouse*. What this means is that, to the extent one actually intends on informing, the value of the choice is not one which is appropriately offset by personal monetary gains from the perpetrator of the relevant wrong. Once one has resolved to intervene, one fails to show due deference to the grounds of one’s choice by allowing personal gains—especially from the perpetrator of the wrong—to interfere with that choice.

To understand how this makes a difference to permissibility consider a case where, as always, threats are no longer at issue. Imagine a large hunting organization is bribing government officials to hunt endangered species. There is too much injustice in the world to say that you have an obligation to get involved. You may certainly, for example, permissibly avoid this issue if you devote much of your time to other just causes. It might even be permissible to base your decision on whether to get involved on certain monetary concerns: It might be too expensive to get involved in this particular issue, so that your time and money are better spent on other issues. But suppose you decide to get involved by disseminating information about the hunters nonetheless. One member of the hunting group takes notice of your activities and offers you money to stop. This is money you *cannot* accept. Now that you are involved in the issue, and you have taken it up as your concern, only certain considerations count as reasons against it—and payment from the perpetrator of the wrong is ruled out. It’s not that the money is ‘dirty’. If this individual had come to you beforehand, and offered it in payment for services or as a gift, you could permissibly accept. It’s the relationship between the money and what it is offered for that is at

issue. This is what makes it a bribe that is both impermissible to request and to independently accept.

The structure of the case is similar when one is choosing whether to become involved in someone else's marriage. Once you take on the task of involving yourself for the sake of the wronged spouse, it is impermissible to be deterred by money offered by the offending spouse. What this means is that if one actually has the relevant grounds for telling the wronged spouse (so that one is not acting cruelly in the telling), it is independently impermissible to seek out such an offer: an announcement to tell unless paid constitutes an impermissible offer (to take an impermissible 'bribe'). So in the third kind of case of marital infidelity we have an explanation of why requesting money for retaining information is impermissible, but it is again different, and much more subtle, explanation than given for the previous two kinds.

Hopefully this discussion helps reveal why it is extremely dangerous to take the case of marital infidelity as the 'standard' case, even though it may be one of the most common actually occurring instances of blackmail. The commonality of case belies a great deal of complexity. It is entirely possible that failures to appreciate the need for subtlety in treating such cases has substantially contributed to the difficulties in giving plausible explanations of the moral mechanics of blackmail.

## 2.2 Punishment, Circularity, and Saving Face

We've now commented on the complete range of statuses that an act of information release could have: merely permissible, impermissible, and obligatory. Even so, there are a few more tricky aspects to cases of threatened information release worth discussion. Sometimes when one threatens to release information, the permissibility of carrying out the threat seems to turn on more than just the harms and benefits that arise were the threat carried out. Moreover, as we'll shortly see, one way this can occur is when the permissibility of carrying out one's threats turns on whether the threat was made in the first place. Both kinds of cases require special commentary. Let's start with the first.

**Case 4A.** Gia steals money from Hana. Hana is unable to prove the wrongdoing, but happens to have some information about Gia that Gia really wouldn't like to see released. Hana announces "Give back the \$N you stole from me or I'll tell everyone your secret".

**Case 4B.** Gia hasn't stolen money from or otherwise wronged Hana, though Hana still happens to have some information about Gia that Gia really wouldn't like to see released. Hana announces "Give me \$N or I'll tell everyone your secret".

It's easy to fill out the details of 4B to make Hana's threat impermissible. But some people have the sense that the threat in 4A is permissible, or at least a lesser wrong. One reason I need to say something about this kind of case is because the only test I've given to check whether a threat to release information

is permissible involves checking the permissibility of the information release itself by somehow weighing the hypothetical harms to the threatened party, and the (appropriate) hypothetical benefits to the party making the threat. But we can imagine that the damage to Gia's reputation is the same in 4A and 4B, and that the gains that Hana gets from acquiring \$N from anyone at all in both cases are at least roughly the same. So what could account for the sense of a strong difference here?

The answer seems to lie in some sense of appropriateness to the use of a threat to get what one is 'rightfully owed'. More generally, it seems less underhanded to use sensitive information to right wrongs in general, regardless of whether one is the victim of the original wrongs or not. We can continue to understand this phenomenon within the simple account of wrongful coercion, I claim, by treating intuitions in 4A as tracking the value in enforcing a just punishment. After all, if we suppose that Gia *doesn't* return the money in 4A, then to the extent it seemed permissible for Hana to threaten the release of information, it likewise seems permissible for Hana to carry out that threat, as a form of retaliation. On versions of 4B where the threat seems impermissible, carrying through the threat likewise seems inappropriate as a retaliatory measure. Viewing the release of information as a form of punishment helps understand why applications of (C) result in different verdicts in cases 4A and 4B, despite similarities in potential personal gains and harms accruing through the fulfillment of the threat: in 4A Hana seems to have a legitimate aim of punishing Gia that is absent in 4B. Further evidence for this view accrues if we make the damage done by the information release disproportionate, and unsuited as punishment for the wrongdoing. Forcing someone to apologize for a snide remark to a coworker by threatening to disclose information that would have them fired, for example, may begin to seem unfairly manipulative.

So far so good. But we can't fully subsume these cases under (C) until we account for one more crucial detail: part of what seems to make the threatened release of information appropriate is the fact that the threat was made in the first place. It's less clear, I think, that Hana can release the information about Gia without any announcement, out of the blue, as a form of 'punishment'. The permissibility of many kinds of sanctions seem to depend on an awareness that the sanction would be enforced in the event of non-compliance.

This immediately raises a circularity worry. The permissibility of the threat, according to (C) depends on the permissibility of what is threatened. But if the permissibility of what is threatened, *qua* punishment, turns on the permissibility of the threat, there seems to be no non-circular way of explaining the permissibility or impermissibility of threats to punish at all. So can we get along only with (C)? Or will we need supplementary principles to break us out of the circular search for justification, in the process threatening the simple and general theory that led us to (C)?

I think the circularity worry is a serious challenge to the application of the simple account because I think the key conditional claim which threatens to generate circularity is correct: *if* the permissibility of retaliation turns on the permissibility of the threat, then vicious circularity will arise in (C)'s applica-

tion. Consequently, to save the application of the simple account we have to deny the antecedent of this conditional. What we've seen so far is that the permissibility of retaliation does seem to turn straightforwardly on a threat *being made*. But note that the troublesome antecedent claims something more. It claims that the permissibility of the threat depends not only on the *existence* of the threat, but its *permissibility*. Even so, this might seem like a very small step. After all, could it really be that the permissibility of performing a particular action depends primarily on whether a threat to perform that action was made, where the threat itself could be impermissible? It can seem odd that this combination of interdependent moral statuses would harmoniously coexist.

Ideally, to defend the utility of (C) we would give a direct counterexample to the claim that the permissibility of executing threats can only ever depend on the existence of a threat but not its permissibility. To do that, we would have to find a case of a threat that is impermissible, but whose execution is permissible. Unfortunately, if (C) is the main way to test for impermissible coercive announcements, this is precisely the kind of case that should be very difficult to find. Instead of going that route, though, I think we can give an 'indirect' counterexample by examining cases where an impermissible threat to do one action can actually render permissible a *different* act of retaliation. Consider the following case:

**Case 5.** Ida is a younger child, old enough to go out with her friends on her own unless, of course, her parents intervene and say otherwise. Ida is thinking about going out tonight with a friend, Jan, with whom she has gone out many times before. Her parents have started to become wary of Jan however. Consider two possible things her parents could announce to her:

- (A) "If you go out with Jan while we're away, you'll be grounded when you get back."
- (B) "If you go out with Jan while we're away, we'll burn every one of your most cherished possessions, one by one, in front of your eyes *and* (to top it off) you'll be grounded."

Let's just suppose that this is a normal case: grounding might be a fair punishment for mild disobedience, but burning cherished possessions is over the top. Since Ida is entitled to go out before her parent's announcement, the simple account of wrongful coercion predicts, I think correctly, that the announcement of 5B is impermissible and that (barring other special explanations of wrongful coercion besides (C)) that of 5A is permissible. But suppose that in 5B Ida throws caution to the wind and goes out with Jan anyway. Suppose further that her parents come to their senses and realize that they can't follow through on their punishment as stated. Instead, when Ida comes back they ground her. Is this permissible? I think so (though of course, the parents probably also owe Ida an apology for their reckless announcement earlier).

This case is of interest because, by stipulation, Ida couldn't be punished in any way for going out with Jan without *some* announcement. But if what I

say about case 5B is right, then she can be punished even after a manifestly impermissible announcement. Why? The case brings out, I think, that when a threat plays a role in legitimating retaliation, it does this by giving ‘fair warning’ that retaliation of a certain *kind* is on the way—perhaps retaliation involving a certain action type or retaliation of a certain severity. To give a warning of this kind just means ensuring that the person threatened (i) realizes that this kind of retaliation may occur and (ii) is given an opportunity to avoid that retaliation. To play these two roles, nothing requires of the warning that it be a *permissible* one. Impermissible warnings—especially those that threaten far more than would be permissible—can still fulfill roles (i) and (ii). That’s why in 5B Ida’s parents can still punish Ida despite having gone overboard in their threat.

Case 5 is admittedly a little contrived, but as I recently noted, these are the kinds of cases we need to use to defend the applicability of (C) if (C) is indeed true and the dominant principle governing impermissible coercion. In any event, now that we have one such case, the circularity worry recently broached is avoided. The legitimating power of warnings doesn’t turn on their own legitimacy. This makes it reasonable to give the description of 4A that I’ve effectively been relying on so far: Hana’s releasing the information about Gia in 4A would be permissible only because Hana antecedently threatened it, but Hana’s threat to release the information is permissible only because it was permissible for Hana to release the information *with the threat to release it having been given*. There is a loose kind of circularity at work here, but it never needs to lead to a circular account of the permissibility of any of Hana’s actions.

This concludes my account of threats to punish through the release of harmful information. But before I conclude this section I’d like to briefly mention one final kind of threat which raises similar issues to that of punishment, and should get a similar treatment: threats which, once made, give the threatener reason to follow through on them in order to ‘save face’.

In normal cases, one only ever makes a threat in the hopes of not having to follow through on it. If the threatener most wanted the outcome where the person threatened does not respond to the threat and the threat is executed, there would be no special need for a contingency announcement in the first place. Accordingly, the efficacy of threats depends on those threatened believing that if push comes to shove, threateners will follow through on their announcement, even if it is challenging for the threatener to do so, or makes the threatener worse off as a consequence. Because one may easily make a threat in a context where one will have to use threats in the future, the need for credibility may provide added reasons for following through on a threat once it is made. Here we have again the potential for the very making of a threat to contribute to the reasons in favor of executing it, and so potentially to contribute to its permissibility. It should be easy to see that the recent account of circularity for threats to punish can be applied here again, in essentially the same way, to explain how this kind of circular relationship can arise consistently with the sole application of (C) to assess permissibility.

## 2.3 Permissible Bluffs

When (C) applies, it tells us that announcements inherit their impermissibility from the impermissibility of what is threatened. Is this always so? A key worry for appeals to (C) are the apparent use of permissible ‘bluffs’: deceptive announcements to do impermissible things in order to achieve worthwhile consequences. Consider:

**Case 6.** Kat is in the process of vandalizing Lya’s property. Lya catches her in the act, but knows that if she calls the authorities they will arrive too late to save her property or to have sufficient evidence to hold Lya responsible. Kat knows Lya’s husband, who is abusive, and knows that if she tells Lya’s husband of the event he will believe her, and beat Lya to punish her. Accordingly, Kat judges it impermissible to actually tell Lya’s husband what she is doing. Nonetheless, she tries to “bluff” Lya into stopping, by announcing that she will tell Lya’s husband if she doesn’t stop vandalizing the property (with no intention of actually carrying the threat out).

Some think that this kind of bluffing is ‘fair game’. Doesn’t (C) predict that it is isn’t?

The answer is actually “no”. (C) is *neutral* on this question. Why? Recall that the justification for applications of (C) came from the simple account of wrongful coercion in interaction with what I called “Permissible Deliberation”:

**Permissible Deliberation:** If one is entitled to deliberate with the possibility of performing action *A* at all, one is entitled to deliberate as if one could perform *A* free of any sanction on its performance that would constitute a prospective wrong.

If we use the *Permissible Deliberation* to arrive at (C), it only applies, as I was careful to state, when we are considering coercive announcements that place constraints on antecedently permissible actions. What we have here is precisely the opposite. Kat is deliberating, or will soon deliberate, about whether or not to continue vandalizing Lya’s property. Not stopping is impermissible, so *Permissible Deliberation* won’t apply to tell us whether Kat is entitled to deliberate with that option without sanction, regardless of whether the sanction is permissible or not.

So in fact my account is so far silent on what is happening in case 6 (and, in fact, even about case 4A from before). Importantly, it can be *extended* in different, but natural ways, to pronounce on the case. One way of extending the account is by accepting additional principles governing impermissible deliberation. For example, the following principle doesn’t seem implausible.

**Impermissible Deliberation:** No person is entitled to deliberate with the option of performing any action *A* that would involve prospectively wronging someone.

The basic motivation for this principle is the fact that we're not entitled to wrong others and we shouldn't be entitled to deliberate with options that we are not entitled to actually perform. Accordingly we're not entitled to deliberate with options that allow us to prospectively wrong others.

If we accept this principle, *and* (C) is the only principle governing the existence of wrongful coercion, then my account will predict that bluffs *never* constitute cases of impermissible coercion provided they only attach sanctions to impermissible behavior. To see why consider case 6 again. Kat is not entitled to deliberate with any outcomes on which, prospectively, Kat intentionally wrongs Lya. This includes any option, regardless of the outcome, on which Kat vandalizes Lya's property. When Lya makes her announcement she is interfering with Kat's deliberation. But, given *impermissible deliberation*, she is not doing so impermissibly: she's attaching a sanction to an action of Kat's which Kat is not entitled to deliberate with *at all*. More particularly, she's attaching a sanction to an action of Kat's which Kat is not entitled to deliberate with *with or without* sanction. Of course this doesn't mean that Lya is entitled to follow through on her threat! Not only is Lya is not allowed to do so, but by *Impermissible Deliberation* Lya isn't even entitled to *countenance* doing so. Interfering with Kat's deliberations by proposing a sanction, and actually following through on that sanction are two very different things. The claim is only that Lya is not *wrongfully coercing* Kat if all Lya does is make an impermissible option in deliberation seem less attractive, however Lya manages to do this.

If we accept *Impermissible Deliberation*, then all cases of bluffs to perform wrongdoing in order to prevent wrongdoing can be treated this way. But even if we do so (which, I should stress, is merely an optional way of extending my account), this doesn't mean that such bluffing never involves any wrongdoing, just that it doesn't involve the special wrongs associated with coercion. After all, bluffing by its nature involves deception. It's entirely possible that only certain kinds of deception are permitted to right wrongs, so that some bluffs, but not others, are exempted from proscriptions against deceptive announcements. Also, deception comes with moral risks. An announced intention invokes trust, in the way that all assertions do, and if an addressee acts on that trust in ways that turns out to be unfairly detrimental to them, the person making the false announcement might be held responsible. Also some false threats, like those to harm innocents, might show an impermissible disregard for other persons regardless of what kind of good is achieved by them. So there are many reasons to think that some bluffs to prevent wrongs might themselves be wrong, even if none are wrongfully coercive.

If one isn't even comfortable with this outcome, this is no problem. The simple account can again be extended in a different direction to treat threats like Kat's as wrongfully coercive. Instead of placing added constraints on what we are free to deliberate with, we can expand the range of threats which count as impermissibly coercive:

**Simple Account**<sup>+</sup>: A coercive announcement is wrongful if it aims to influence an agent's deliberation over whether or not to do *X* by threaten-

ing to do something impermissible depending on whether or not the agent does  $X$ .

*Simple Account*<sup>+</sup> involves a strengthening of the combination of *Simple Account* and *Permissible Deliberation*. The latter two principles rely on entitlements to deliberation, and maintain that in cases where one has such entitlements, others cannot influence your choice by attaching impermissible sanctions to the choice you deliberate with. This leaves open that perhaps one can attach impermissible sanctions to impermissible actions permissibly, as *impermissible deliberation* would ensure. By contrast, *Simple Account*<sup>+</sup> dictates that whether or not an agent has an entitlement to deliberate with some choice is actually irrelevant. One can never attach impermissible sanctions to any actions permissibly, even if the actions are themselves impermissible. This would ensure that bluffs of the kind that Lya make are not just impermissible, but impermissibly *coercive*.

The neutrality of simple account is one of its virtues. It can be coherently extended in different ways to cope with corresponding intuitions about coercive bluffs. Moreover we only need the bare structure of the simple account to cope with a vast and diverse range of coercive threats of information release.

### 3 Broader Implications of the Account

I haven't addressed every complication that could arise in explaining what makes instances of blackmail impermissible, but the range of cases I've examined hopefully foster the sense that careful attention to neglected kinds of moral disregard like cruelty can successfully resolve the paradoxes of blackmail and give us an account of the wrongs involved in blackmail through the *Simple Account* and (C) that is simple, general, and quite powerful. In what space remains I want to turn from defending the account itself to briefly examine its broader implications. First I'll discuss what we can learn from my defense here for theories of coercion more generally. Then I'll make some brief remarks about the legal versions of the puzzles about blackmail.

#### 3.1 Blackmail and the Simple Account of Impermissible Coercion

Blackmail as I've been considering it in this paper involves (i) using the threat of the harmful release of information in (ii) an act of impermissible coercion. In accounting for the wrong done in blackmail it is natural to focus on either of these two features to try to explain what is particular to its operation. It is not uncommon to focus on the fact that blackmail involves sensitive information, and to try to use this as a way of unraveling the difficulties and paradoxes that arise in understanding the prohibition on blackmail. So, for example, Murphy (1980) tries to argue that the reason that blackmail is illegal is to protect privacy. If blackmail weren't illegal, Murphy argues, it would ultimately create a harmful market for private information. Owens (1988) argues that blackmail is

impermissible in part because it involves an ‘unrenderable’ service: blackmailers offer to keep certain information from getting out, but prototypically no one person can really ensure this, leaving the person blackmailed at the mercy of further blackmailers offering the same ‘good’.

In contrast, the account I’ve been giving focuses entirely on the fact that blackmail involves coercive threats. On this account there is nothing about information (beyond the minimal fact that its release can be harmful) or privacy (beyond the minimal fact that it is one among many values) which figures in an account of what makes blackmail wrong. Such an account has two key virtues. First, it avoids the special problem, besetting accounts that focus on the status of information, of explaining what is so special about information that would make our exchanges in it, as opposed to more conventional forms of property, governed by special mores or regulations. Second, an account which successfully subsumes the wrongs involved in blackmail under a more general account of wrongful coercion strengthens both. The latter account is now seen to have an attractive simplicity and generality which was easy to overlook because of complicating factors arising in particular cases. Moreover, the tools we develop in coping with special cases of blackmail are now free to be reused in accounting other instances of wrongful coercion more generally.

After all the *moral* paradoxes of blackmail are, on reflection, not obviously particular to blackmail at all. It is easy to raise parallel puzzles for other kinds of coercion. Consider

**Case 7.** Moe is a big time real estate developer and is trying to buy up land to make a new golf course. He needs the land from Nora’s property to do so, but Nora has no interest in selling. There is a standing law which dictates that if one of two neighbors wants to build a fence between neighboring properties, the first neighbor may force the second to share half the costs. The fencing around Nora’s land would actually be quite expensive for Nora, and Nora would find the fence a nuisance as it would block the beautiful views she presently has from her property. Using this information, Moe announces “If you don’t sell to me, I swear I’ll buy up all the property around you and sue to ensure you pay half the costs of the fencing. There, does that make selling worth your while?”

Moe’s threat is morally impermissible. But isn’t it morally permissible for Moe to buy the land, and to build a fence around it if he wants? And if, before any threat, Nora comes to Moe to sell her land because the surrounding land is being bought up for a golf course, can’t Moe permissibly accept the offer? This case, and others besides, present exactly the same kinds of puzzles as the traditional *moral* paradoxes of blackmail. And because they present it for the very same reasons, all the work done within the framework afforded by the simply account of wrongful coercion applies to enable us to resolve these puzzles as well.

Now, though these cases are identical from the *moral* perspective, from a legal one they are different. Moe’s threat in this case may well be sanctioned by

the law. Why the legal asymmetry between these coercive uses of property and the coercive uses of information more specifically? This is a very good question. Let's briefly see why.

### 3.2 The Legality of Blackmail

I began this paper by noting that the paradoxes of blackmail come in both moral and legal varieties. So far I've only focused on the moral versions of these puzzles and, in part for this reason, had to forego a discussion of the vast legal literature on the topic. Without delving too deeply into the details of the legal debates, though, I believe we can say something general about the importance that unraveling the moral paradoxes might have for them.

Broadly, resolutions of the legal paradoxes of blackmail fall into two categories: *internal* and *external*.<sup>12</sup> Internal theories look to something inherent to the act of blackmail itself—the threat, the exchange of goods, carrying out the threat—to explain the legal puzzle. External theories look to the harmful external consequences of allowing blackmail to transpire unfettered.

Two examples of external theories are given by Nozick and Epstein.<sup>13</sup> Nozick claims that blackmail is prohibited because it is 'unproductive'. Normally purchasers of goods are better off for the transaction, even if the prices are exploitative. By contrast the victim of blackmail would have been better off had the blackmailer not existed. The problem is not with any violation of rights, but a kind of inefficiency. To take another example, Epstein claims the principal problem with blackmail is that legalizing it would lead to systematic cases of fraud against third parties—those who are owed the information which is withheld in a successful instance of blackmail. Again, there is nothing inherently wrong with the acts of blackmail, just their 'downstream' effects.

External theories often result in inadequate extensional classifications of illegal blackmail (e.g. Epstein's argument only seems to apply to cases where the third party is legally entitled to the information, which covers only a small number of cases of blackmail). But they obviously suffer more systematically from an inability to explain why blackmail is a violation of the *rights of the person blackmailed*. Why in the transaction, is the blackmailer, for example, the only one punished?

Internal theories try to sidestep these difficulties by looking to interpersonal violations of rights. But to the extent they try to make sense of the existing laws, they tend to fall prey to the paradoxes of blackmail quite directly. If one takes on board that releasing the information is protected under the law, it becomes extremely challenging to give a convincing explanation of why threats to release the information shouldn't be protected as well.

The difficulties are compounded by the fact that formulations and interpretations of blackmail law suffer from failure to appreciate the complexities of the

---

<sup>12</sup>The terminology is from the helpful survey of the legal literature on blackmail in Wertheimer (1987) pp.92–103, but the broad form of classification is also used to great critical effect in Lindgren (1984). I closely follow both able discussions here.

<sup>13</sup>Nozick (1974) pp. 84–86, Epstein (1983).

paradoxes of blackmail. As Lindgren puts it:

The paradox of blackmail is not merely an abstract philosophical question. A failure to understand or resolve the paradox has spawned a body of law that is in disarray—statutes that do not adequately describe the crime and court opinions that are poorly reasoned or just plain wrong.<sup>14</sup>

The resolution to the moral paradoxes provides some insight into the vexed state of the legal literature. What the resolution of the moral puzzles teaches us is that *if* the illegality of blackmail is grounded primarily in moral concerns, then the law is giving distinct treatments to two kinds of behavior which should on a legal par. The moral basis for banning certain threats to release information is the very same moral basis for banning the release of that information (provided, of course, that the motivational structure of the threatener is held constant, since this was a key ingredient in assessing moral permissibility in both cases).

This really only leaves us with two options for addressing the legal puzzles. First, we can admit that moral concerns ground the illegality of blackmail and go in for a form of legal revisionism: either the forms of information release that are presently legal should, in the special cases in which cruelty is involved, be made illegal, or what is presently classified as illegal forms blackmail should be made legal. This option puts my view close to that espoused by libertarians like Block & Gordon (1986), but for very different reasons. It's no special libertarian stance which provides any impetus towards legalizing blackmail. The present view merely says that the cases might need to be treated on a par, if moral grounds for illegality are all that is at issue. Settling on legality may simply be the most prudent way to resolve in favor of parallel treatment.

There is a second option though. We can allow that other concerns are at work in the legal setting which make a differential treatment reasonable. An obvious concern here might be enforceability. Forbidding certain kinds of information release that are morally impermissible because cruel will require tests for such cruelty. As such they will inevitably involve highly speculative inquiries into the values and intentions of individuals releasing sensitive information. Perhaps this problem is avoided in cases of blackmail because one's values are made more transparent when one actively seeks compensation for retaining information. If so, perhaps we can capitalize on differences in the availability of information about motives to draw a distinction between threats to release information and acts of information release themselves. I'm not entirely sure whether this distinction can be safely drawn and enforced, or whether other distinctions might make for an appropriate differential legal treatment. Accordingly, I'll be content for now to leave this as a special problem for future legal research.

---

<sup>14</sup>Lindgren (1984) p.676.

## References

- W. Block & D. Gordon (1986). 'Blackmail, Extortion and Free Speech: A Reply to Posner, Epstein, Nozick and Lindgren'. *Loy. LAL Rev.* **19**:37–54.
- W. Block, et al. (2000). 'The Second Paradox of Blackmail'. *Business Ethics Quarterly* **10**(3):593–622.
- M. Clark (1994). 'There is no Paradox of Blackmail'. *Analysis* **54**(1):54–61.
- S. DeLong (1993). 'Blackmailers, Bribe Takers, and the Second Paradox'. *University of Pennsylvania Law Review* **141**(5):1663–1693.
- R. Epstein (1983). 'Blackmail, Inc.'. *The University of Chicago Law Review* pp. 553–566.
- H. Evans (1990). 'Why Blackmail Should be Banned'. *Philosophy* **65**:89–94.
- J. Feinberg (1988). *The Moral Limits of the Criminal Law: Harmless Wrongdoing*, vol. 4. Oxford University Press.
- D. Ginsburg & P. Shechtman (1992). 'Blackmail: An economic analysis of the law'. *U. Pa. L. Rev.* **141**:1849.
- J. Lindgren (1984). 'Unraveling the Paradox of Blackmail'. *Columbia Law Review* **84**(3):670–717.
- J. Lindgren (1993). 'Blackmail: An Afterword'. *University of Pennsylvania Law Review* **141**(5):1975–1989.
- J. Murphy (1980). 'Blackmail: a preliminary inquiry'. *The Monist* **63**(2):156–171.
- R. Nozick (1974). *Anarchy, state, and utopia*. Basic Books.
- D. Owens (1988). 'Should Blackmail be Banned?'. *Philosophy* **63**:501–514.
- T. M. Scanlon (2008). *Moral dimensions: Permissibility, Meaning, Blame*. Belknap Press.
- A. Wertheimer (1987). *Coercion*. Princeton University Press.
- G. Williams (1954). 'Blackmail'. *Criminal Law Review* **79**:79–92, 162–72, 240–46.