

Chapter 6

On Coerced Promises

Eric Chwang

Abstract

It is less rational to rely on a coerced promise than it is to rely either on a coerced (non-promissory) action or on a non-coerced promise. In this paper I defend a two-part explanation for that intuitive judgment. First, coerced promises carry no moral force. That is, they do not generate additional moral reasons, above and beyond whatever reasons already exist, for the promiser to fulfill her promise. This undercuts the idea that the general moral obligation to keep one's promises can be used to defend the reliability of coerced promises—that general obligation does not apply to coerced promises. I argue for this thesis by, first, explaining why the importance we attach to the activity of promising cannot extend to coerced promises and, second, refuting four recent alleged counterexamples to the thesis. The second part of the explanation is that coerced promises carry no predictive force either. That is, they do not communicate reliable intentions to fulfill the promise, even independently of moral motivations. This undercuts the idea that the promissory intentions can be used to defend the reliability of coerced promises—the general connection between promising and intending does not apply to coerced promises. I defend this thesis by arguing that, while it is possible to coerce promissory intentions, rational agents may be unable to promise sincerely when threatened, and, further, coerced promises typically give no evidential support that the promiser intends to do as she says.

I. THE PROBLEM

Suppose a mugger jumps out of the bushes, points a gun at his mark, and insists, “Your money or your life!” The mark convinces him she has no money on her person, so the mail-in mugger says, “Promise me you’ll mail me some

money tomorrow; otherwise I'll kill you now." If the mark acquiesces then she makes a coerced promise. The mail-in mugger might think that he is quite clever, to have extracted a promise of money when he could not extract instantaneous money. However, coercing the promise of future money is in an important sense grasping at straws. After all, it is clearly not as profitable as coercing immediate money. A typical mugger who coerces his mark for \$100 can expect more money than the mail-in mugger coercing a promise for the same amount via post. It is also more rational to rely on an uncoerced promise for money than on a coerced one. An Oxfam representative, for example, might accost me on the street and implore, "Please mail some money to Oxfam tomorrow!" If I acquiesce and thereby promise, the Oxfam representative can be more sure that I will keep my promise than the mail-in mugger can be that his mark will keep hers.

Why these differences? One potential explanation is that it is conceptually impossible to coerce a promise. This is wrong—Margaret Gilbert has argued decisively that it is possible to coerce a promise.¹ I will not pause to review Gilbert's argument or defend her conclusion here, due to space constraints. Instead, I will merely assume that it is possible to coerce a promise, and I will argue that two other facts explain the sense that attempting to coerce a promise is akin to grasping at straws. First, coerced promises carry no moral force. That is, they do not generate additional moral reasons, above and beyond whatever reasons already exist, for the promiser to fulfill her promise. However, this negative moral thesis is not the entire explanation. After all, even if coerced promises exert no moral force, it might still be rational to rely on them, because people use promissory language to communicate their intentions. Thus, I will also argue for a second thesis: that coerced promises typically carry no predictive force either. That is, they do not reliably create or indicate extra intentions to fulfill the promise, above and beyond whatever intentions the promiser already had to fulfill it.

Two crucial points are in order before continuing. First, I claim only that reliance on a coerced promise is irrational (or even more cautiously, that reliance on a coerced promise is less rational than reliance on a coerced act or on a noncoerced promise), not that the attempt to coerce it is irrational. It may well be rational for the mail-in mugger to attempt to coerce a promise, for example, even if I am right that the resulting promise has neither moral nor predictive force. That is because his unfortunate mark may be ignorant or skeptical of this essay's central thesis. If the attempt to coerce the promise costs little, then it may well be prudentially rational for the mail-in mugger to go ahead—the net expected pay-off of attempting it may be greater than that of not bothering. My goal is explain why, in spite of this, the attempt to coerce a promise is still a much less reliable means of ensuring performance of the promised act than either coercing the act directly or else eliciting the

promise noncoercively. Attempting to coerce a promise is akin to grasping at straws in a way that neither attempting to coerce a nonpromissory act nor eliciting the promise without coercion are.

Second, coerced promises will typically be artificial in at least one respect. If, for example, the mail-in-mugger can credibly threaten his mark with harm after he was to have received her money, then he is arguably still coercing an action, rather than coercing a promise of future action. It is just a more protracted mugging for money that takes place over a span of several days, rather than a routine mugging that lasts at most a few minutes: “(Promise me you’ll) mail me some money tomorrow; otherwise I’ll find you and kill you next week.” He may even continue to use the word “promise,” as in the parenthetical addition in the previous sentence, but such use would be confused. To ensure that our examples of coerced promises are not confused in this very immediate way, we must build into them the assumption that the coercer cannot produce a credible threat after the promise has been given. For example, we might suppose that the mail-in mugger is abroad and is leaving for his native country tomorrow, never to return.

II. MORAL FORCE

Coerced promises exert no moral force, which means that they do not provide any additional moral reason to fulfill the promise. Of course, this is compatible with the claim that one ought to or perhaps even is obligated to perform the promised act in question. In this section I will explain why we ought to deny moral force to coerced promises. In the next section, I will examine four recently proposed alleged counterexamples to the thesis that coerced promises carry no moral force.²

The Value of Promises

Why do we value promises? By this I mean to ask not why an individual might value some particular token promise, but rather something like why we as a society of rational and communicative agents value the practice of making and keeping promises. The answer to this latter question strikes me as very obvious and straight-forward. The reason we value the ability to make a promise is that we want to give promisers the moral power to create and then assume obligations on their own. For example, if I promise to help you move, then I thereby, all by myself, create and then assume a moral obligation to help you move. Certainly, there are other ways I can acquire the obligation to help you move, without my promise, but the reason promising is important and useful is that it gives me a way to assume that obligation voluntarily, all by myself.³ Now, if

my promise is coerced, then in the relevant sense I am no longer acting alone and voluntarily in assuming the promissory obligation. Rather, it is more accurate to say that it is being foisted on me. Thus the essential moral kernel we want to capture in recognizing the importance of promising is incompatible with giving moral force to coerced promises. The point of granting moral power to promissory activity is to give the power to promisers to let them incur obligations when they choose, all by themselves. And that does not happen when one is coerced into promising. In the remainder of this section I will make six further points that help clarify this idea.

Against Counterbalancing

Consider a counterbalancing picture, according to which promises always carry moral force even when they are coerced, but this moral force is counterbalanced by the immorality of coercion. Sometimes the immorality of coercion suffices to outweigh the moral force of the coerced promise, and sometimes not, depending on various factors. Such a picture may seem attractive, but it is incorrect.

One reason it is incorrect is that it is part of the moral principle about promissory moral force that the promise in question cannot be coerced. In other words, the principle “promises have moral force” is really shorthand or elliptical for the more informative principle, “uncoerced promises have moral force.” It is not as if promises are like Ross’s prima facie duties, so that there are multiple principles at play, one telling us that promises always have moral force (without exception), and others restricting the range of that first principle, so that we end up with a downstream rule according to which coerced promises can have various force, depending on how weighty the various principles are in particular cases. Rather, the one principle about promises says that uncoerced promises have moral force. (And this principle is almost certainly elliptical or shorthand for some even more complex principle, for example one ruling out deceived promises.) That is because, again, the point of promissory activity is to allow people to create and then assume obligations all by themselves. That coerced promises carry no moral force is built into this idea.

I have just argued that the counterbalancing view is incorrect because it is inconsistent with the best explanation of the importance of promissory activity. Another reason it is incorrect, though, is that it generates counterintuitive results independently of my larger thesis. For it is not in general true that if my means of acquiring your promise involves a coercive element then your promise thereby loses moral force. I might, for example, employ coercion on a third party to acquire your promise, in which case it retains all its original moral force. For example, suppose Amber wants to offer an employment

contract to Beatrice, but she is too lazy to bring it to Beatrice in person. Amber might coerce an innocent passerby, Carrie, to bring the contract over: “Carrie, bring this contract to Beatrice, or else I’ll shoot you!” Carrie, under threat of death, acquiesces. Beatrice knows that the employment proposal is between her and Amber, and that Carrie was merely the coerced messenger, so the proposal itself is not coerced, though coercive means were used to make it. In this case, the mere fact that the promise (contract) was elicited immorally and via a coercive step does not thereby reduce its moral force. Beatrice is not allowed to renege on the resulting promise just because Carrie had to bring over the contract at gunpoint. (Imagine, to make this vivid, that Amber and Beatrice are white American landowners before the Civil War, and Carrie is Amber’s slave.) Of course this is compatible with the thought that Beatrice should voice her indignation at Amber’s wrongful coercion, but reneging on the resulting promise is not automatically the best or only way to do so.

Consent as Analogy

Consider the analogy between promising and consenting. We generally think that consent can change what would otherwise be an impermissible act into a permissible one. For example, if I do not consent to it, then you are not allowed to drill in my teeth, but if I do then you might be. I said “might be” because, of course, there are exceptions, such as if the consent was coerced. For example, if a parent coerces her child to consent to tonsillectomy, that consent thereby carries no moral force. This is compatible with the claim that the surgeon still ought to operate, and it is compatible with the claim that the child ought to have consented (without being coerced). But the coerced consent itself should have no bearing on the surgeon’s deliberations about whether he ought to proceed. Indeed, in spite of a massive bioethics literature on consent, no strand of thought in that literature attempts to argue that coerced consent can be valid; coerced consent is uncontroversially recognized to be invalid, without exception. The reason for this is that the practice of recognizing consent allows people to create and then give, all by themselves, moral permissions to others. If I consent to your drilling in my teeth, then I create a moral permission to drill and give that permission to you. Of course, there are other ways in which you might acquire that permission, without my having a say in the matter, but the point of consent is to allow consenters to be able to create and give moral permissions voluntarily, all by themselves.

Coercion versus Inducement

Compare coerced promises with induced ones. I will discuss in more detail later the difference between coercion and inducement, but for now we can

say that inducements are welcome but coercion is not. For example, I would rather not be confronted by a mugger waving a gun in my face and saying “Your money or your life.” In contrast, I welcome a potential employer waving an unsigned contract in my face and saying “I’ll pay you a hefty salary to come work for me.”

Both coercion and inducement are forms of outside influence, of changing the payoffs of an agent’s actions. However, only coerced promises are unduly influenced in the way that compromises promissory moral force. That is because a coercer is manipulating the situation surrounding the agent’s choice in an unwelcome way, but an inducer does so in a welcome way. A coerced agent thus is precluded from going down the path she otherwise would have chosen (keeping both her money and her life), and she is thereby less responsible for her choice between the remaining alternatives. In contrast, an induced agent is also influenced, but in a way she welcomes. Thus, in the relevant sense the induced agent still made her choice voluntarily, without undue outside influence, so that it still appropriate to assign responsibility to the induced agent for making the choices she made. “I couldn’t help it; he waved a gun in my face” is a legitimate excuse that absolves responsibility for one’s actions—such as handing over one’s wallet or taking a job—in a way that “I couldn’t help it; he waved an unsigned contract in my face” is not.

Coercion versus Force

We should also distinguish coerced promises from what we might call forced promises.⁴ A coerced promise, roughly, requires an intentional agent to coerce, whereas a forced promise is made on the basis of unfortunate circumstances in nature. For example, if you put a gun to my head and threaten to shoot unless I give you money, you are attempting to coerce me to give you money. However, if through no fault of anyone’s I need sugar quickly (I am diabetic), and you are a nearby convenience store proprietor, then I may be forced to give you money in exchange for some orange juice. You are not attempting to coerce me, though circumstances may force me into give you money regardless.

Now of course this “coerced” versus “forced” terminology is arbitrary and stipulative, but the distinction it demarcates is morally significant. Coerced promises never have moral force, though forced promises often do. That is, if I am coerced into promising to mail you some money tomorrow (you have a gun to my head), then the resulting promise does not create an additional moral reason for me to carry out the exchange. In contrast, if my diabetic condition forces me to promise to mail you some money tomorrow for your orange juice today, then by promising I do thereby generate an additional moral reason to mail money to you.⁵

Three-party Coerced Promises

What about a three-party case? Suppose, for example, that Arnold coerces Bill into promising to help Charlie move, where Charlie is ignorant of the coercion. Does Bill's promise generate a moral reason for him to help Charlie move? No, because it was coerced. Of course, this is compatible with the thesis that Charlie is justified in relying on Bill's promise and justified in being indignant that Bill reneged. Yet here, as usual, coercion acts as an excusing condition: because Bill was coerced, he is excused from responsibility for reneging on the resulting promise. This is analogous to nonpromissory coercion. Alice coerces Barbara to break Connie's finger, where Connie is ignorant of the coercion. Connie is justified in relying on Barbara not to break her finger and justified in being indignant at her broken finger, but Barbara is not responsible for it.

Now, of course there is the important difference that in the promissory case what is coerced is the promise-making, not the promise-breaking. Thus one might suspect that an agent can be coerced into making a promise, for which she is thereby not responsible, and yet she might still be held responsible for breaking it. This is incorrect: if one is not responsible for making a promise then one is likewise not responsible for breaking it. We can preserve this intuitive principle by distinguishing between the moral force generated by promises and that generated by expectations.

A pair of contrasting examples will help explain this distinction. Suppose first that Jones has the justified yet false belief that Smith promised to help him move. Smith knows all this, and he knows that helping Jones move will be much less costly than reneging. Plausibly, Smith has a moral reason to help Jones move, in spite of not having promised, and that is because Jones's reliance generates a moral reason, even though Smith is not responsible for this reliance. Now consider a contrasting example, where Jane knows that Samantha promised to help her move, yet Samantha (faultlessly) does not remember doing so. If Samantha knows that Jane is relying on her to help, then that generates a moral reason for Samantha to help, the same reason as for Smith. However, if Samantha were then to discover the truth—that she did in fact promise—then she would realize she has an additional moral reason to help, not shared with Smith: the moral force of a genuine promise.⁶

This distinction also explains the intuition that Bill ought to help Charlie move, even if Arnold coerced Bill into promising to do so. Bill's coerced promise has no moral force, but this is compatible with Charlie's (faultless) reliance having moral force. After all, Bill's unfortunate situation is analogous to Smith's. Both, if they are sensitive moral agents, should feel an unease because someone is faultlessly relying on them for help. And those expectations are enough to generate moral reasons to help. Yet in both cases the

reliance has been foisted on them, so they are not responsible for it. In Smith's case, that is because no promise has, in fact, been made; in Bill's case it is because his promise was coerced, and therefore without moral force.

The Enforceability Objection

Promissory practice is valuable because it gives promisers the power to create obligations all by themselves. One might object to this by wondering whether there might be other reasons to value promissory practice, reasons that can license moral force in coerced promises. The short answer is that no other such reasons exist, and the slightly longer answer is that the burden of proof is now on the defender of coerced promises to come up with one. After all, my explanation—that we want to give people the power to incur obligations, all by themselves—seems sufficient to explain what we find important in promissory activity, so that any additional explanation would add nothing. Still, perhaps something is missing. In this subsection I consider what seems the most promising alleged omission: it might sometimes be in the victim's interest for coerced promises to be enforceable; otherwise, the coercer might do something even more horrible than merely coerce a promise. For example, the mail-in mugger might kill his mark if he cannot get a guarantee of future money. Thus, the inability to enforce coerced promises can lead to worse consequences for their victims.⁷

Several points can be made against this objection. First, it concedes that the consequences are often better if we refuse to enforce coerced promises. In that case, for example, the would-be mugger might just walk away on discovering that the mark has no cash on hand. Further, even if the consequences in some isolated cases would be better if we enforced coerced promises in those cases, we might still be best off with an exception-less rule against enforcing any coerced promises. This is because, once we admit that we will sometimes enforce coerced promises, there will be an incentive for would-be coercers to raise the stakes in their interactions, so that they too can be covered by the exception clause. In other words, indirect consequences via reputation effects matter too.⁸

A second worry is that this objection applies most naturally to contracts, not to promises. Mere promises are not enforceable in courts, regardless of whether they are coerced. Thus, the mail-in mugger might be disappointed to learn that he cannot appeal to the legal system to enforce his mark's coerced promise, but in just the same way my neighbor might be disappointed to learn that he cannot appeal to the legal system to enforce my uncoerced promise to help him move. My concern in this essay is the moral force of promises, not the legal validity of contracts.

A third problem also relies on the distinction between ethics and the law. Enforceability is at bottom a legal matter, whether the law should enforce

various promises (or better, contracts). Moral force is of course a moral matter. Now, one way to justify a legal rule, say of enforcing certain coerced promises, is by appealing to the analogous moral rule, here that those coerced promises have moral force. But the absence of this kind of straightforward inference from moral to legal is compatible with another method of justifying the legal rule, via appeal to protecting potential victims. Thus, it is possible that the objector and I are both right: coerced promises have no moral force, and yet the law should sometimes enforce them anyway.

A final problem is that enforceable promises no longer generate the puzzle I want to address. That is because an enforceable coerced promise is in the relevant sense equivalent to a protracted coerced action. For example, if the mail-in mugger can appeal to the courts to ensure that he gets his money within a week, then he can produce a credible threat of harm—namely legal punishment—at the time when he is supposed to receive the money. Obviously, in that case he is perfectly sensible in attempting to coerce a promise, in which case the example is no longer interesting. I am interested in explaining why, for example, a promissory mugging is unreliable when the mugger cannot produce any credible threat at the time that the money changes hands, including even the threat of legal sanction. The phenomenon I want to explain would not arise in a world in which all promises are enforceable.

III. ALLEGED COUNTEREXAMPLES

In this section I consider four recently alleged examples of coerced promises with moral force. I argue that each suffers from the same two defects. First, they are most plausibly construed as coerced actions rather than as coerced promises, and once they are modified to be genuine cases of coerced promises they become too artificial to be applicable to the real-world concerns that might have motivated us to countenance moral force in coerced promises to begin with. Second, they blur the distinction between threats and offers, which while delicate is extremely important in the conceptual analysis of coercion. The examples are genuinely coercive only if they are made on the basis of threats, in which case the resulting promise carries no moral force, and they carry genuine moral force only if they are made on the basis of offers, in which case they are not coerced.

Surrender

I begin with wartime surrender.⁹ My discussion of this example will be by far the lengthiest, because the same points I develop here apply straightforwardly to the other examples as well. Suppose Winning is winning a war

against Losing, and that Winning makes the following proposal to Losing: “Promise to lay down arms and never pick them up again, and we will stop fighting; otherwise, we will annihilate you.” Losing acquiesces. This may seem like a morally forceful coerced promise, but in fact it is not.

Surrender as Coerced Action versus Coerced Promise

First, typical cases of surrender are best interpreted as coerced action rather than as coerced promises. Recall the caveat I mentioned at the end of section 1: if Winning is able to produce credible threats against Losing after the time of surrender, that surrender is essentially a coerced action—here the action being the laying down of arms and refraining from picking them up again—rather than a coerced promise of future action. This is so even if promissory language is confusedly invoked.

This point is important because one might mistakenly think that the only way we can make sense of the significance and legitimacy of the actual practice of surrender in the real world is by countenancing the moral force of coerced promises. That is false; in actual cases of surrender the victor is never so foolish as to relinquish the ability to produce credible threats against the loser after surrender. In other words, the best explanation for the legitimacy of real-world surrender proposals need not invoke morally forceful coerced promises. In the real world, victorious nations do not relinquish the ability to make credible threats against their surrendering foes.

Threats and Offers

Of course, we can construct artificial cases where Winning is not able to produce a credible threat against Losing after the time of surrender. For example, we could just stipulate that Winning, perhaps foolishly, will melt down its weapons once the surrender is signed. However, now a second problem for the surrender example emerges, namely that the example straddles the distinction between threats and offers. If the surrender proposal is a threat, then while it is certainly coercive it also carries no moral force. On the other hand, if the surrender proposal is an offer, then while it carries moral force it is thereby no longer coercive.

The difference between threats and offers is roughly that if I propose to make you better off then my proposal is an offer, whereas if I propose to make you worse off then it is a threat. For example, proposing to shoot you if you do not hand over your money makes you worse off; proposing to exchange my goods for your money makes you better off (even if you

reject the proposal, on the understanding that your rejection carries no penalty). If we add the innocuous principle that there is an essential, if rough, connection between proposals you welcome and proposals that make you better off, then we can connect this analysis of threats and offers with the earlier analysis of coercion and inducement: threats coerce, and offers induce.

Of course, questions regarding what makes you better or worse off are essentially comparative. Robert Nozick, in his seminal article on coercion, notes that two kinds of comparative baselines exist. Moral baselines ask whether you are worse or better off than you should be, and predictive baselines ask whether you are worse or better off than you would otherwise be.¹⁰ Now, we do not have to settle the best way to think about baselines, because my conclusions follow regardless. That is fortunate, for the proper way to demarcate threats from offers is disputed.¹¹ The important point for my purposes is not that some one particular way of distinguishing is best, but only that we distinguish.

Surrender as Threat

Suppose first that Winning's surrender proposal is a threat. An easy way to do this is to use a moral baseline and assume that Winning does not have the moral right to annihilate Losing. After all, if we use a predictive baseline then real world surrender proposals seem always offers, because in the natural course of events that would otherwise transpire the victor would annihilate the loser. Suppose, then, that Winning has no right to annihilate Losing, yet Losing surrenders and then breaks its promise and begins to fight again. If Winning were to complain that Losing has broken its promise not to pick up arms again, how might Losing respond? Of course, one possibility is that Losing does not care about morality at all. But another option is open, one that is both within the bounds of morality and perfectly reasonable: their earlier promise was made on the threat of annihilation, and Winning did not have the right to make this threat. Therefore, the promise is morally bankrupt, because coerced promises have no moral force.

An example may help persuade. Let Winning be Nazi Germany, and Losing be France. Nazi Germany is not morally permitted to annihilate France, and it coerces France to surrender via the threat of annihilation otherwise. Now suppose French citizens are debating whether to start a resistance against the occupation. If one citizen were to say, "we should not start a resistance, because we promised that we would not when we surrendered," it would be perfectly natural for another to respond, "but that surrender carries no moral force, because we were (wrongfully) threatened with annihilation if we didn't."¹² The first citizen, if he is to continue this debate, must

say more about why France ought to respect the surrender agreement even though it was coerced.

In other words, if we assume that Winning is not morally justified in continuing to fight to the point of annihilating Losing, then surrender is not a persuasive counterexample to the thesis that coercive promises have no moral force. The reason it is not persuasive is that it is perfectly reasonable to interpret the example in a way friendly to the thesis that it allegedly contradicts. A persuasive counterexample would have to be an uncontroversial case that is obviously a morally binding coerced promise. The example of surrender made on the basis of threat is too controversial to be persuasive.

I should stress this point about dialectic and the burden of proof. I have argued that there is a general presumption in favor of the thesis that coerced promises never carry moral force. Any counterexample to that thesis must be less controversial than the thesis it is being used to refute, if it is to be effective. For, in the face of any controversial counterexample, it is always open for the defender of my thesis to reject the counterexample. Here is an analogy. Suppose I want to argue that one can know something without knowing that one knows it. And suppose my candidate example is my knowledge that I am not a brain in a vat: I intend to prove that I do not know that I know that I am not a brain in a vat. Clearly, this is a bad argumentative strategy, because my assumption is shaky—it is contentious whether I even know that I am not a brain in a vat in the first place. Of course, contentiousness is compatible with truth: it might very well be true that I know that I am not a brain in a vat. But, dialectically, if I want to convince my interlocutors I should not begin by presupposing something so controversial, to wit that I know that I am not a brain in a vat.

The White Flag Objection Objection: What if Losing surrenders after it has displayed a conventionally understood symbolic gesture of surrender, such as waving a white flag or falling to its knees? The moral power of the white flag is uncontroversially accepted, after all.¹³ So suppose Losing initiated the surrender proceedings by waving a white flag. Surely in this case if Losing were to pick up arms again it would be acting immorally; we would judge Losing to be dastardly and dishonorable if it waved the white flag and then turned around and attacked again after Winning let down its guard.

We must be careful when thinking about white flag scenarios. Recall that we are presupposing a coercive context in which Losing waves the white flag: Winning is not permitted to annihilate Losing yet is threatening to do so regardless. In such a scenario, it is perfectly reasonable to infer that if Losing waves the white flag it is coerced into doing so. After all, as Nozick has reminded us, a threatener can coerce without making an explicit proposal.¹⁴ In particular, it can be clear from the circumstances that

Winning will not initiate its own surrender proposal but will instead continue to fight until one of two things happens: either Losing is annihilated or else it waves the white flag and then surrenders. In such a situation Losing is clearly coerced into waving the white flag, but then intuitively the resulting surrender agreement still carries no moral force. Suitably modifying the France and Nazi Germany example should make this point clear.

Granted, white flag intuitions are much stronger in cases where the winning country is permitted to annihilate the losing country. For example, if Losing started an unprovoked, unjustified war of aggression on Winning, but now is facing annihilation, then we think that Losing would be dastardly or dishonorable to renege on its surrender agreement if it initiates that agreement by waving the white flag. However, in this kind of case the surrender proposal that results is an offer rather than a threat, which brings us to the other horn of my initial dilemma.

Surrender as Offer

Suppose, then, that Winning's surrender proposal is an offer. One way this might be true is if the distinction between threats and offers is made on the basis of a predictive baseline. For we can stipulate that, as seems typically the case in the ¹⁵real-world, if Winning were not to make the surrender proposal it would just annihilate Losing instead. However, another way Winning's surrender proposal might be an offer is if we use moral baselines and then assume that Winning is morally permitted to annihilate Losing—an example might be (a suitably cleaned-up variant of) the Allies proposing surrender to Japan in World War II. Now, suppose Winning, perhaps in response to Losing waving the white flag, makes the following proposal: "Look, we are permitted to annihilate you, but we are going to offer you a deal: we will refrain from annihilating you—which, again, would be perfectly within our moral rights—but you have to promise never to pick up arms again." This proposal is not coercive, because it is an offer rather than a threat. It is an offer rather than a threat because it proposes a course of events that is better for Losing than the relevant baseline scenario, no matter whether that baseline is predictive or moral: surrender is better than annihilation. Therefore, even if Losing promises on the basis of an offer, and even if that promise carries moral force, the promise was induced rather than coerced.

Surrender as Coercive Offer Objection: my argument thus far assumes that only threats can coerce, whereas perhaps offers can coerce as well. Notably, David Zimmerman and Joel Feinberg have each independently argued that

offers can coerce.¹⁶ In this subsection, I will argue for two points. First, Zimmerman and Feinberg are wrong. Second, the conceptual possibility of coercive offers is irrelevant for my thesis as originally intended anyway, in which case this objection is just a request for unnecessary precision.

Zimmerman says that an offer is coercive if the offeree both prefers it to the preproposal situation and prefers to move from the actual preproposal situation to some alternative preproposal situation that is feasible, and that the offerer has actively prevented the offeree from obtaining.¹⁷ A surrender offer is likely to count as coercive on Zimmerman's analysis because the loser might prefer surrender to continued fighting and yet might also prefer a return to something like the prehostility state of affairs to both. The prehostility state of affairs (or something approximating it) is feasible, and the winner's military campaign is actively blocking the losers from obtaining it. Thus, surrender offers are coercive.

Zimmerman's conceptual analysis is incorrect, however. Situations which Zimmerman would describe as Al making a coercive offer to Bob are better described merely as situations where both (i) Al (wrongfully) harms Bob by actively preventing Bob from obtaining something Bob wants and (ii) Al exploits Bob by proposing an offer that is not as good as it should have been. Bob has no additional grievance of coercion in Zimmerman-cases, and we do better to reserve coercion-talk for cases where Al threatens Bob. After all, suppose (i) and (ii) both hold in some offer situation, and then Al makes the further move of putting a gun to Bob's head and saying "Okay, it's no longer an offer; now it's a threat: do what I want, or I shoot you." Zimmerman has no more moral vocabulary to describe the moral worsening of the situation that has just occurred. He could, of course, invent a term to cover this more serious moral offense, but then the obvious reply is that by "coercion" we meant to be discussing the referent of this other term all along.

According to Feinberg, an offer is coercive if it proposes a choice between two evils, such that one of the two evils is greatly preferred to the other.¹⁸ For example, if I am happily employed in Colorado, but then offered an even better job in New York, that offer is not coercive. In contrast, suppose my job in Colorado is miserable (evil), and I am offered another job in New York that is also miserable, though significantly better than my Colorado one. In that case the offer is coercive. Arguably, surrender offers are coercive on this account as well. The choice between surrender and annihilation is the choice between two absolute evils, yet surrender is greatly preferred to annihilation.

Feinberg's analysis is defective on two counts. First, it wrongly presupposes that there is a universal metric by which we can determine whether a state of affairs is absolutely good or evil, full stop. States of affairs can be better or worse than one another, but this is an essentially comparative

judgment. The only sense we can make of absolute yet nonarbitrary judgments of the goodness or badness of states of affairs is by neglecting to articulate our implicit and perhaps most contextually salient comparison class. For example, when I say that my job in Colorado is wonderful, I am implicitly comparing it to many other worse and particularly salient jobs that I might have had instead. But it makes no sense to repudiate all implicit comparison classes and then to say that my job in Colorado is wonderful, full stop.

The other problem with Feinberg's analysis is that it robs coercion claims of their intended significance. One of the most important reasons we care about coercion is that it reduces responsibility: if I am coerced into doing something, then I am not morally responsible for having done it. However, Feinberg readily admits that one is still morally responsible for accepting a coercive offer.¹⁹ For example, if you offer me a job in New York, then I retain full responsibility for my choice to take it, even if both the new job and my old one were evil on absolute terms. In contrast, if you put a gun to my head and threaten to kill me if I do not take it, then I am no longer responsible for having taken it. The most theoretically elegant way of expressing this difference is as the difference between inducement and coercion.

I do not have the space for a full discussion of the inadequacies of theories which countenance coercive offers. However, even if I am wrong, three further points can be made. First, even if I am wrong it is at least controversial whether coercive offers are possible. Besides Nozick, Alan Wertheimer is perhaps the most notable scholar on coercion, having produced the first and most influential book-length philosophical treatment of the subject, and he also rejects Zimmerman's and Feinberg's arguments, concluding that if we care about coercion for moral and legal purposes then we ought to rule out the possibility of coercive offers.²⁰ But, to repeat, dialectically persuasive counterexamples cannot rest on controversial assumptions.

Second, I am perfectly happy to restrict my thesis so that it is compatible with the existence of coercive offers. When I said that coerced promises have no moral force, what I really meant to say was that coerced promises *made on the basis of threats* have no moral force. The italicized clause is vacuous, but it will provide clarification for those who (mistakenly) think that offers can coerce. And the more cautious thesis, that coerced promises made on the basis of threats carry no moral force, is surely the kernel of truth in the repudiation of the moral force of coerced promises. Taking a cue from Feinberg, we can say that agents retain moral responsibility if they choose to accept a coercive offer. For example, Losing retains moral responsibility for accepting Winning's coercive surrender offer. After all, in the important and relevant sense, Losing was still able to make the promise all by itself, without

undue outside influence. Even coercive offers are welcome, so Losing is still responsible for its choice.

Third, recall my original motivation: to explain why it is foolish to rely on a promise that was coerced. In light of the distinction between threats and offers, my motivation can be restated more precisely: to explain why it is foolish to rely on a promise that was coerced *on the basis of a threat*. As before, the italicized addition is vacuous, but it may be helpful for those who (mistakenly) think that offers can coerce. There is nothing especially problematic about relying on a promise that was coerced on the basis of an offer.

Contrasting two examples should make that last point clear. Let us suppose in both examples that my job in Colorado is an absolute evil. Now, in *Miserable New York*, a recruiter proposes to let me come work for him in New York, as long as I promise to mail him some money. (He will never see me again, but he left me a mailing address.) This job in New York, while significantly better than my very evil Colorado job, is still evil on absolute terms. Contrast that with *Wonderful California*, where a different recruiter makes a noncoercive offer: the California recruiter approaches me and offers an absolutely good job in California in exchange for the promise of future money. *Wonderful California* may raise interesting issues about, for example, how self-interested agents can guarantee reciprocity in exchange, but it does not have even a whiff of grasping at straws, which is my explanatory target. And, for all intents and purposes, *Miserable New York* is the same—it raises no further interesting issues. That is, neither the *Wonderful California* nor the *Miserable New York* recruiters would be irrational to rely on my promise in the way that the mail-in mugger would be. Recruiters can rationally expect most people to follow through on their offer-based promises, but muggers should not expect most (rational) people to follow through on their threat-based promises. Thus, even if the phenomenon I want to explain is not captured precisely by the simple language of coercion, it is still adequately captured by the slightly more complex language of threat-based coercion. And in the case of offer-based coercion, the irrationality phenomenon that I want to explain does not even arise.

Shotgun Wedding

A second alleged example of a coerced promise with moral force is the shotgun wedding.²¹ Pa holds a shotgun to Ben's head and tells him that he needs to promise to marry to Emily, whom (it is common knowledge that) Ben has impregnated. As before, we need to make some artificial additions to the example in order for it to be plausible that Pa is coercing the promise of marriage rather than merely coercing the wedding (or marriage) itself. Otherwise the example is most naturally interpreted as one where promissory

language is confusedly used to describe a protracted coerced action. That is important because, as before, we do not need to license moral force in coerced promises in order to explain the legitimacy of real-world shotgun weddings. In a real-world shotgun wedding, what is coerced—if anything is coerced at all—is not a promise, but rather a prolonged action.

But suppose, for the sake of furthering the argument that (it is common knowledge that) Pa has a terminal disease and will die before the wedding day. Is this a morally forceful coerced promise? No. As with the surrender example, there is a dilemma depending on whether Pa is morally permitted to shoot Ben for having impregnated Emily. First suppose no. In this case the promise of marriage is indeed coerced, because it was made on the basis of a threat. However, it is perfectly natural to say that the coerced promise is therefore morally impotent. Of course, this is compatible with it being the case that Ben ought to marry Emily and indeed that Ben ought to promise to marry to Emily. The claim here is only that the coerced promise generates no moral force of its own, to add to previously extant considerations.

Perhaps an analogy will help. Suppose Johnny (intentionally and wrongfully) steps on some other child's sand castle. Of course, Johnny ought to apologize, but suppose Johnny's mother points a gun at Johnny's head and says "Promise to apologize, or I'll shoot you." (It is common knowledge that, tragically, Johnny's mother will be whisked away to a foreign country after the promise but before the apology, never to return or be in any contact with Johnny ever again.) Johnny acquiesces. Does Johnny's coerced promise exert any moral force? No. Yes, Johnny ought to apologize, and perhaps Johnny even ought to promise his mother that he will apologize without her having to resort to threats to acquire that promise. But if Johnny promises and then refuses to apologize, he has committed only one wrong (in addition to the wrong of stepping on the sand castle), namely the wrong of not apologizing. He has not also committed the further wrong of breaking a promise. And that is because his promise was coerced (not because coerced promises are impossible, which is false). As with the surrender case, we can account fully for our moral intuitions about this case without having to add an alleged extra wrong of a broken coerced promise.

Likewise in the shotgun wedding. On the assumption that Pa is not permitted to shoot Ben with the shotgun in retaliation for impregnating Emily, Ben's coerced promise of marriage carries no moral force. And this is compatible with the claims that Ben ought to marry Emily and even that Ben ought to promise Pa to marry Emily without Pa's having to threaten him first.

Now suppose that Pa is morally permitted to shoot Ben for having impregnated Emily.²² In this case, the shotgun wedding proposal is an offer, not a threat. Pa says, "Look, I am allowed to shoot you with this shotgun in response to what you did to my daughter, but I propose instead that you marry her,

which is better for you than being shot.” Offers do not coerce, or, if they do, not in the way that is of interest here. Thus, the shotgun wedding promise either carries no moral force or else is not coercive.

Parental Discipline

David Owens suggests that a child’s promise to behave, elicited via parental coercion, has moral force.²³ This is incorrect. Again, we must first eliminate the worry that parents typically are able to produce a credible threat after the time at which the child must carry out her promised action. Otherwise this case, like those that preceded it, is best interpreted as one where the parents confusedly use promissory language to describe a protracted coerced action. And, as before, the contortions necessary to eliminate this worry will render the resulting case artificial—because the parent will have to lose contact with her child after the time of the promise—and therefore irrelevant to real-world motivations for acknowledging morally forceful coerced promises.

But even after we waive that worry, the same dilemma that arose for surrender and shotgun weddings remains. Either the parent is morally permitted to carry out her proposed punishment, or not. If not, then her coerced promise has no moral force—Johnny and the threat of being shot in the head for trampling the sand castle illustrates this point.

If, on the other hand, the parent is morally permitted to carry out her proposed punishment, then the parent is making an offer, which is not coercive. Suppose, to modify the sand castle example, that the mother’s punishment is ten minutes of quiet time (instead of being shot in the head), and suppose that the mother is morally permitted to wield this punishment in response to Johnny stepping on the sand castle. In that case, the mother’s proposal, “Apologize, or else you get ten minutes of quiet time” is again an offer, because it proposes a course of action—the apology— that is expected to be preferable to the morally permitted and otherwise expected option of ten minutes of quiet time.

Judges and Criminals

The last alleged example of a coerced promise with moral force also comes from Owens; it is the custom of some judges to use the threat of punishment in order to get criminals to promise to reform.²⁴ It is not merely that the judge coerces reform via the threat of punishment. No, the practice in question is that the judge also coerces the promise of reform. Once elicited this coerced promise allegedly carries its own moral force, so that a criminal who fails to reform is then guilty of two wrongs (failing to reform and breaking a promise) rather than just one.

As before, we must take care to ensure that the examples are clear cases of coerced promises rather than of coerced actions. In particular, it is quite natural to interpret judges as confused if they insist not just that the criminal behave but also that the criminal promise to behave. Really, they are incorrectly using promissory language to coerce an action (good behavior). After all, the legal system will still be able to produce credible threats on the criminal after the time of his promise. We can of course construct fanciful yet coherent examples where the legal system cannot produce a credible threat after the criminal is set free, in which case this promissory language would not be confused. But, as before, such artificial examples bear too little resemblance to the real world to be able to conclude that common, everyday, real-world considerations commit us to the moral force of coerced promises.

And again, the same dilemma arises. Either the court is morally permitted to carry out its threat, or it is not. If it is not, then the proposal is coercive, because a threat, but it carries no moral force. For example, suppose the criminal shoplifted, and the threat is the death penalty. The proposal, “Promise not to shoplift, or else we’ll execute you” is certainly coercive, but the promise that results is morally impotent. That is, even if the criminal ought to refrain from shoplifting, and even if he ought to promise to refrain (without being threatened first), the fact that the criminal’s promise was coerced means the promise does not add any moral force to the moral demand that he refrain. If the shoplifter is coerced into the promise to refrain—via the threat of the death penalty—and then goes on to shoplift, he may be guilty of having shoplifted, but he is no more blameworthy for having also broken the coerced promise.

If, on the other hand, the court is morally permitted to carry out its threat, then the proposal is best construed as an offer, and is therefore not coercive. For example, if the legal system is morally permitted to incarcerate the criminal for a week, then the proposal “Promise to behave, or else we will incarcerate you for a week” is an offer, and therefore not coercive, even if the resulting promise has moral force.

IV. PRACTICAL FORCE

In this section, I will argue that coerced promises typically carry no practical force. That is, coerced promissory behavior is not good evidence of intention to fulfill the promise. In a typical case, if I say that I promise to do something, then you can conclude that I intend to do it. And in a typical case where it is common knowledge that I intend to do something, you can raise your epistemic credence that I will do it. Might this work for coerced promises? If, for

example, the mark gives a coerced promise to mail some money, does this give the mail-in mugger a good reason to think that the mark will mail the money, independent of moral concerns?

We must disentangle some questions. First, we can ask about the connection between promising and intending. One way the mail-in mugger might be confused is in thinking that a genuine promise to do something automatically carries with it the genuine intention to do that thing. That is false, because insincere promises are still promises. Indeed, one can promise to do something without intending to undertake an obligation to do that thing as well, and insincere promises are again good examples.²⁵ At this point one might further conclude that, even if it is possible to coerce a promise, it is not possible to coerce promissory intentions (such as the intention to fulfill the promise or the intention to undertake the obligation to fulfill a promise). However, this further conclusion would be too hasty.

Possibility of Coerced Intentions

Is it possible to coerce intentions? One easy answer to this question appeals to an analogy to beliefs. Plausibly, it is not possible to choose to believe something when one has strong evidence to the contrary, in which case it is likewise impossible to coerce someone into believing it. Analogously, perhaps it is not possible to coerce intentions, in the same way that it is not possible to coerce beliefs. But this answer is too easy because, even if doxastic voluntarism is false, it is not clear that intentions are analogous to beliefs in the relevant respects. Yes they are both mental states, but that is insufficient to motivate the analogy between them.

Another easy answer to this question is that, yes, it is possible to coerce intentions because, in general, if I coerce you into X-ing, and X-ing requires intending to X, then I have also coerced you into intending to X. For example, if I coerce you into making a cucumber and avocado sandwich for me, and your making a cucumber and avocado sandwich for me entails your intending to make a cucumber and avocado sandwich for me, then I have also coerced you into intending to make a cucumber and avocado sandwich for me. This opposing answer is also too easy, for two reasons.

First, it may not even apply to the relevant promissory intentions. Promissory intentions can include the intention to fulfill the promise and also the intention to undertake the obligation to fulfill the promise. The easy answer now under consideration entails that I can coerce you to intend to make a cucumber and avocado sandwich by coercing you into making it—the intention comes along for free, so to speak. But it is less clear whether this move will work for the intention to undertake an obligation to make the

cucumber and avocado sandwich. What action could I possibly coerce which would bring along that obligation-intention for free? If I put a gun to your head and say, “Undertake an obligation to make a cucumber and avocado sandwich for me, or else I’ll shoot,” you may mouth agreement just to get the gun out of your face, but it would be quite implausible to suppose that I can thereby generate obligations in that manner, let alone the intention to undertake obligations.²⁶

Second, the most interesting cases of coerced promises are ones where the promise is coerced precisely because it is not possible to coerce the promised action. Recall that the mail-in mugger attempts to coerce the promise of future money precisely because the mark has no money on her person to give.

A more fruitful approach to our question than either of the two previous easy answers begins by noting that one can form intentions voluntarily in cases where it may be impossible to form analogous beliefs voluntarily. For example, I can choose whether to intend to eat a cucumber and avocado sandwich for lunch, though I cannot choose whether to believe that a cucumber and avocado sandwich is on my desk now. Given that I can voluntarily form intentions in a way that I cannot voluntarily form beliefs, the suggestion that intentions can be coerced even though beliefs cannot is more promising. For, if I can choose whether to do something (give you money, form an intention), then it seems I can be coerced into doing that thing (give you money, form an intention). After all, if I can choose to do something then I can choose on the basis of reasons, and attempts at coercion merely provide a certain type of reason—namely conditional threats—that are not special in any relevant way.

Now, even if it is granted that we can choose our intentions on the basis of reasons, including conditional threats, there seems to be something especially problematic about coerced promises. That is because in a typical case of coerced promises the promisee will have no good reason (moral, prudential, or otherwise) to follow through on her intention, once the time comes to fulfill the promise. For example, once the mark is no longer threatened by her mail-in mugger, she has no good reason to follow through on her promise by mailing him money. Then the question arises whether it is possible for her to form an intention to do something that she can foresee in advance she will have no good reason to do.

I have just described some essential features of Gregory Kavka’s toxin puzzle.²⁷ A toxin will cause you to be in pain temporarily if you ingest it. A wealthy person proposes to give you a lot of money at midnight tonight if, at that time, you intend to ingest the toxin the following morning. All that is required for your receiving the money is that you intend tonight to ingest tomorrow, not that you actually ingest tomorrow. Indeed, the

money will be transferred to you before you have to ingest. Is it possible to intend to ingest, when you know that ingesting will be bad for you, though not so bad that you would prefer abstaining to acquiring the money?

There is an extensive literature on the toxin puzzle, but thankfully our question here is simpler than the hardest questions one might ask about that puzzle. The most interesting questions about the toxin puzzle relate to what is possible for rational agents, but none of us are perfectly rational in the sense required for standard decision theory. Thus, we want to know whether it is conceptually possible to coerce an intention in a potentially irrational agent. In particular, even if it is impossible for a perfectly rational person to intend to do something she knows she will have no good reason to do, it still is at least conceptually possible that a mail-in mugger might coerce an less-than-perfectly rational mark into the relevant promissory intentions.

Futility of Coerced Intentions

Thus, it is at least conceptually possible to coerce intentions. Still, even if it is conceptually possible to be coerced into forming an intention, lingering questions remain. One question, as we saw at the end of the previous section, is whether it is possible to coerce a perfectly rational agent into intending to do something that she knows she has no good reason to do. Returning to the mail-in mugger, is it possible for a perfectly rational mark to intend to mail the money, if she foresees that she will have no good reason (moral, prudential, or otherwise) to do so when the time comes?

On the analogous toxin puzzle, Kavka thought that a rational agent will not ingest the toxin, and so cannot intend to. Others disagree, arguing that because it is in the agent's interest to intend, that gives her a (defeasible) reason to inculcate the disposition to ingest when that time comes, so that it can be rational to intend and indeed, come morning, even to ingest as well.²⁸

As I said earlier, this dispute over the toxin puzzle is difficult, and I have nothing novel to say about it. I mention it because commitments on the toxin puzzle carry over to the analogous case of the mail-in mugger. I am content to conclude conditionally that if Kavka is right then it is impossible for a rational mark to form promissory intentions in the mail-in mugger case. In that case, it is futile to attempt to coerce a promise from a (known to be) rational mark.

More can be said, however, for there is another, less contentious, argument for futility. In particular, the potential mail-in mugger faces another practical problem: How will he ever know that his mark has formed the relevant

promissory intentions? In other words, even if he can coerce a promise, how will he know that the mark's promise is sincere? The analogous question is glossed over in discussions of the toxin puzzle—where it is assumed that the player's intentions are transparent—for the sake of other more theoretical concerns. However, our interest here is practical, so the question poses a legitimate concern.

The worry is that observable promissory behavior is much less trustworthy if it is coerced. In a typical uncoerced case, if I say I promise to mail some money to Oxfam, that promissory performative counts as evidence that I have formed the relevant promissory intentions. However, such performative evidence is not nearly so strong in a coerced case. For a coerced performative indicating that I intend to do something is not (good) evidence that I so intend. In other words, if you coerce me into saying I promise to do something, then my saying that I promise is no longer good evidence that I intend to do it, and no longer good evidence that I intend to undertake the obligation to do it.

This also suggests that attempting to coerce promises will typically be futile. Not merely because rational people cannot intend to do what they know they will not have any good reason to do, which is disputed, but rather (or also) because an attempted coercer will typically be unable to acquire sufficient evidence that he has succeeded in coercing the relevant intention.

I say “typically” because this is not always the case. The victim may be horribly bad at bluffing, for example, so that if she were to say “I promise to mail some money to you” without the requisite intention, she would send an involuntary observable cue—a facial tic, perhaps—to the coercer that she lacks the requisite intention. Then the sophisticated coercer might try to coerce her victim into the promissory performative, not because that suffices for her to have promised and promises automatically generate moral or predictive force, but rather because he can watch her performance for clues as to her intentions. This is atypical. In a typical case, the coercer's access to his victim's intentions is mediated by the victim's voluntary behavior, and such behavior is no longer a reliable indicator of intentions when the victim has an external incentive to engage in promissory behavior.²⁹

Note that, in principle, similar remarks apply to induced promises as to coerced promises. That is, while induced promises are conceptually possible, there is a worry about whether the promiser's performative is good evidence that she has the requisite intention. For example, supposing I know that you will give me \$1,000 today if I merely promise to house-sit for you next month, I have an incentive to engage in a promissory performative, even when I do not intend to fulfill my promise.³⁰

However, there are two important asymmetries between typical cases of coerced versus induced promises. First, considerations of reputation differ. That is, we are typically more concerned with maintaining a good reputation with people who try to induce us into promises than we are with people who try to coerce us into promises. For example, I am more concerned to develop a good reputation amongst potential house-sitting employers than I am to develop a good reputation amongst potential mail-in muggers. Second, induced promises do, but coerced promises do not, have moral force. Therefore, it is typically more wrong for someone to renege on an induced promise than on a coerced one.³¹

V. CONCLUSION

I started with the thought that it is irrational to rely on coerced promises, or more precisely that it is less rational to rely on coerced promises than it is to rely on uncoerced promises or on coerced actions. The explanation for this is two-fold. First, coerced promises have no moral force. Second, coerced promises typically have no practical force as well. In other words, a person attempting to coerce a promise has no good reason, whether based on what the promisee ought to do or on what she intends to do, to expect in typical cases that a (rational) coerced promisee will fulfill her promise. Of course, I do not expect this essay to convince the people who might actually try to coerce promises; my audience is philosophers, not would-be promise-coercers. Rather, exploration of this topic is intended to shed light on conceptual and moral issues that arise from the interrelated notions of promises, coercion, moral force, and intentions.

ACKNOWLEDGMENTS

Thanks to participants at two talks at the University of Colorado at Boulder's Center for Values and Social Policy and to participants at the Rice University conference Understanding Promises and Agreements for valuable feedback. Special thanks to David Boonin, Chris Heathwood, Hanoch Sheinman, and Alan Wertheimer for comments on earlier drafts. Thanks to Emma Kobil for some last-minute proofreading. And finally, thanks to Amber Arnold and Barrett Emerick, whose innocent questions about the possibility of self-promises exposed a loose thread in a much earlier draft, the persistent unraveling of which necessitated that this final version be very different, basically rewritten from scratch—thanks a lot, you two.

NOTES

1. See Gilbert (1993). Gilbert decisively rebuts two arguments. The Obligation Argument claims that promises by definition obligate, but coerced promises never do. This argument is flawed because promises do not always obligate. The Voluntariness Argument claims that promises are by definition voluntary but coerced promises are not. This argument is flawed because it equivocates on two senses of “voluntary.” Strictly speaking, Gilbert’s arguments concern the possibility of coerced agreements, and agreements are not the same as promises. (If the upshot of her essay is right, agreements are not even mere promise-pairs.) But her arguments about the possibility of coerced agreements apply straightforwardly to coerced promises too. See Gilbert (2006): 228.

2. See, most notably, Gilbert (1993), Deigh (2002), and Owens (2007).

3. What I say here is not intended to be novel; it is intended to be a statement of common sense. It is therefore similar to what many others have said about promises. For example, Owens (2007) argues that we care about promises not only because they give us predictive information about what the promiser is likely to do but also because promises generate rights of authority in the promisee. And Searle (2001) insists that promises are agent-created reasons. I merely combine these two insights.

4. See, for example, Kavka (1986): 396.

5. What about a case where the convenience store proprietor knows about my diabetic condition, and so jacks up the price of orange juice? He is not attempting to coerce me, though he is attempting to exploit me. Do exploitative promises carry moral force? For example, assuming commercial surrogacy contracts exploit, does a desperate commercial surrogate’s promise to relinquish the baby have moral force? This issue is interesting, yet tangential to my concerns. The intuitive idea that promises give us the power to create and assume obligations all by ourselves does not dictate an answer to this question about exploitative promises. For, even if the desperate surrogate’s exploited promise was made in the relevant sense voluntarily or all by herself, there may be legitimate reasons to curb or restrict the power to create and assume voluntary obligations.

6. A third contrasting kind of case will make the distinction even more clear: one can make a promise on which (it is common knowledge that) the promisee will not rely. In such cases, the promise can still generate moral reasons to fulfill even if there is no reliance on the fulfilling.

7. For a statement of an argument along these lines, see Bar-Gill and Ben-Shahar (2005).

8. This is just the sort of consideration that many find persuasive on the question of whether to negotiate in hostage situations. Even if the consequences of negotiation are sometimes favorable, it might be better to adopt an exceptionless rule against negotiation, to inculcate a hardline reputation which eliminates the incentive for future hostage-taking.

9. Deigh (2002) defends surrender as the only type of coerced promise that has moral force.

10. Nozick (1969): 447–449.

11. For some nice discussion and a full bibliography, see Anderson (2008).

12. At this point, assume that the surrender proposal is initiated by Nazi Germany, rather than by France. I will soon discuss the scenario where France initiates the surrender proceedings by, e.g., waving a white flag.

13. Deigh's discussion of the moral importance of conventional signs such as waving the white flag or falling to one's knees at Deigh (2002): 490 suggests this.

14. Nozick (1969), 444.

15. can the hyphen be deleted here? usually used only when adjective.

16. Zimmerman (1981); Feinberg (1986): ch. 24.

17. Zimmerman (1981): 132-133.

18. Feinberg (1986): 235.

19. Feinberg (1986): 248.

20. Wertheimer (1987): chapter 12. Wertheimer's discussion is more nuanced and complex than my simplified summary, but not in a way that is harmful to my overall point.

21. Gilbert uses this as an example of a binding coerced agreement in Gilbert (1993): 702 and following. It also works as a coerced promise, if we assume (it is common knowledge that) the marriage proposal will be accepted.

22. As with the surrender case, we might also suppose that we ought to use predictive baselines, and that had Pa not made the surrender proposal he very likely would have shot Ben. The argument using moral baselines is more vivid, so I will continue to make my points with that implicit assumption.

23. See Owens (2007): 298. Owens misdescribes the scenario as one of parental warning, but warnings are distinct from threats (and from offers), and warnings do not coerce. (And unlike my analogous insistence that offers do not coerce, the thesis that warnings do not coerce is undisputed.) See Nozick (1969): 453-458 for more on this distinction. This error is easy to remedy, though of course "parental coercion" sounds less like what a loving parent should do than does "parental warning."

24. Owens (2007): 298.

25. This latter question is relevant because an influential thesis about promising is that a promise to do something communicates, not the intention to do that thing, but rather the intention to undertake the obligation to do that thing. See, for example, Raz (1977).

26. Note that even if Raz is right to say that promises communicate the intention to undertake obligations, it does not follow that if I attempt to coerce you into undertaking obligations, I am thereby attempting to coerce you into promising.

27. Kavka (1983). Of course, the toxin puzzle is an offer, not a threat.

28. See, for example, Gauthier (1998) and Harman (1998).

29. More cautiously, the victim has that external incentive when dealing with a naïve coercer, one who wrongly thinks that a coerced performative is still good evidence of intentions. If it is common knowledge between victim and coercer that performatives in the context of attempted coercion are not good evidence of intentions, then the victim of course no longer has any reason to engage in such performatives. Perhaps in this second-stage common knowledge situation, engaging in performatives once again becomes good evidence of intentions, in which case the victim once again has a strong external incentive to engage in them. This leads us to a third-stage common knowledge situation, where the coercer and victim both know all the above.

Then the victim once again no longer has any reason to engage in performatives. But that leads to a fourth stage of common knowledge. And so on. As long as engaging in performatives incurs some cost, this game may lack a pure solution.

30. Considerations related to Kavka's toxin puzzle also apply to induced promises, as this example should make clear.

31. If the moral importance of promise-keeping reduces to reputation effects then I have just stated the same asymmetry twice.

32. pls. give full name

REFERENCES

- Anderson, Scott. (2008). "Coercion." In Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*, fall 2008 ed., <<http://plato.stanford.edu/archives/fall2008/entries/coercion/>>.
- Bar-Gill, Oren. and Ben-Shahar, Omri. (2005). "Credible Coercion." *Texas Law Review* 83, no. 2 (February): 717-780.
- Deigh, John. (2002). "Promises under Fire." *Ethics* 112, no. 3 (April 2002): 483-506.
- Feinberg, Joel. (1986). *Harm to Self*, vol. 3 of *The Moral Limits of the Criminal Law*. New York: Oxford University Press.
- Gauthier, David. (1998). "Rethinking the Toxin Puzzle." In Jules L. Coleman and Christopher W. Morris, eds., *Rational Commitment and Social Justice: Essays for Gregory Kavka*. Cambridge: Cambridge University Press, 47-58.
- Gilbert, Margaret. (1993). "Agreements, Coercion, and Obligation." *Ethics* 103, no. 4 (July 1993): 679-706.
- . (2006). *A Theory of Political Obligation*. Oxford: Oxford University Press.
- Harman, Gilbert. (1998). "The Toxin Puzzle." In Jules L. Coleman and Christopher W. Morris, eds., *Rational Commitment and Social Justice: Essays for Gregory Kavka*. Cambridge: Cambridge University Press, 84-89.
- Kavka, Gregory. (1983). "The Toxin Puzzle." *Analysis* 43: 33-36.
- . (1986). *Hobbesian Moral and Political Theory*. Princeton: Princeton University Press.
- Owens, David. (2007). "Duress, Deception, and the Validity of a Promise." *Mind* 116 no. 462 (April): 293-315.
- Nozick, Robert. (1969). "Coercion." in ³²Morgenbesser et al., eds., *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*. New York: St. Martin's Press, 440-472.
- Raz, Joseph. (1977). "Promises and Obligations." In P. M. S. Hacker, ed., *Law, Morality and Society: Essays in Honour of H. L. A. Hart*. Oxford: Clarendon Press.
- Searle, John. (2001). *Rationality in Action*. Cambridge, MA: MIT Press.
- Wertheimer, Alan. (1987). *Coercion*. Princeton: Princeton University Press.
- Zimmerman, David. (1981). "Coercive Wage Offers." *Philosophy and Public Affairs* 10, no. 2 (spring): 121-145.