

Contracts under Coercion

Should one Keep an Agreement with a Robber?

Abstract: It is nearly universally accepted, both in law and in morals, that agreements entered into under coercion are not binding. Thus you are not obligated to keep, and the law will not enforce, a contract you undertake to save your life at gunpoint. What is the theory behind this invalidation of coerced contracts? Most explanations see coercion as vitiating the voluntariness of the agreement, in either a psychological or a moral sense, and conclude that coerced agreements do not truly reflect the victim's will. Alternatively, scholars argue that coercive contracts should not be enforced because that would encourage the use of coercion. In this Article I begin by arguing that the appeal to involuntariness to explain the invalidity of coerced contracts is not compelling. The deterrence approach rests on an empirical truth, but it is problematic for other reasons. I then revisit the common wisdom that agreements entered into under coercion are non-binding. I suggest that a more compelling argument for the conventional wisdom can be discovered, ironically, through a careful study of the only philosopher who rejects the conventional wisdom about coerced agreements, namely Thomas Hobbes. Using Hobbes' discussion as a point of departure, I offer a Hobbesian-style argument for the invalidation of contracts entered into under coercion. While the Hobbesian perspective for which I argue ultimately defends the conventional wisdom about the enforceability of coerced contracts, it does so in a way that is less sweeping and more narrowly drawn than traditional theories have proposed.

I. INTRODUCTION: THE PROBLEM

Imagine you are jogging along a deserted street in the evening when an arm hooks around your neck from behind and a deep voice growls, “Your money or your life.” You feel something hard press into your back, which you take to be a gun. You explain you were out for a jog with no money—not even an ATM card. You have plenty of money in the bank and are willing to return the next day with \$5,000 cash, provided that your assailant releases you now. He doubts your sincerity, but you explain, “This deal is in my interest. After all, the alternative is considerably worse for me. If you believe me you will release me, and if you do not, you will kill me, assuming that you are rational, since no murder to your credit and a good deal of cash is superior to no money and a high likelihood of arrest for a serious felony.” Therefore, I have every reason to convince you of my sincerity to abide by the agreement if there is any way for me to do this. Your assailant considers this argument. “So you see,” you continue,” “we will both be better off and neither of us worse off under the deal I propose than if you proceed to kill me now.” Should the robber believe you? Assume for the moment that the robber is rational, that you are rational, and that each of you has knowledge of the other’s rationality. In this case, it seems unlikely you will be able to convince the robber that you regard it as in your interest to keep your promise if it is not. The “common knowledge of rationality” that holds between you effectively makes bluffing promises highly unlikely to succeed. From this it follows that you will be able to convince the robber you will actually return with the money only if it is, indeed, rational for you to return with the money. So the question now is: *would* it be?

It would be possible to *make* it rational for you to return with the money if there were some sort of penalty you could set up to ensure your future performance. If, for example, he had some way of following up should you fail to appear the next day, matters would be different.

But alas, he cannot verify any information you might give him about yourself, as you have left all personal identification at home. To your mutual frustration, the situation seems hopeless. But finally an idea dawns on you which you quickly propose to your assailant: you offer to write a *contract* in which you will commit to returning with the promised sum the next day in exchange for your immediate release. You explain that you are a professor of contract law and are reasonably adept at drafting such documents, and you assure him that you will spare no pains to make sure the contract is correctly written. Satisfaction of the statute of frauds, unambiguous terms, a liquidated damages provision, and a binding arbitration clause—nothing will be left out! But to your dismay, your assailant turns out to be none other than . . . a law student! And he now points out that none of these measures could possibly help. He reminds you of a little impediment to your plan known as the *duress* defense,¹ which will be a bar to enforcement of this kind of contract in every court in the land. He cocks his gun and prepares to shoot.

“Wait!” you shout. “What if we agree to *suspend* the duress defense as part of our contract?” You explain that since in this case *both* parties would prefer to eliminate the defense, allowing the suspension is surely a win-win proposition. Furthermore, you argue, the duress defense is for the benefit of the victim, and here you—the victim—are proposing yourself to eliminate it! Wouldn’t any sensible and compassionate judge honor an agreement between the parties to suspend the defense under *these* circumstances? Your assailant, however, is once again skeptical. Any clause suspending the duress defense would *itself* be written under duress, he points out, given that it is written at the very moment at which he is threatening to shoot. It might have been different if the agreement to eliminate the duress defense had been entered into at a prior time, before you were threatened with force. But even then, he says, courts would be reluctant. The duress defense, he reminds you, is not a mere *default* rule, but a *mandatory* rule,²

and for good reason: contracts under coercion are not supposed to be binding under *any* circumstances, for *whatever* reason. Individuals are not allowed to exempt themselves from this rule just because there are particular occasions on which the rule works out badly for them. You are almost certainly out of luck, he explains.

As the above discussion makes clear, you and your assailant face the following paradoxical situation. You and he would both like to enter into an enforceable agreement in which he relinquishes his hold on you now, in exchange for your promise to return the next day with the money. The agreement, considered at the present moment, would benefit both, relative to a no-agreement baseline. But no matter how fervently you both insist that your agreement is mutually beneficial, no matter how much you insist that there is a “meeting of the minds”³ between you, no matter how sincere your intentions and how great the loss of disallowing such agreements, no court in any jurisdiction would enforce it. Moreover, this result is not the least bit controversial. The proposition that contracts entered into under duress are unenforceable is as solid a proposition of black letter law as ever there was.⁴ Like their legal counterparts, philosophers are similarly unanimous in the view that commitments made under coercion entail no moral obligation.⁵ Can the confidence with which courts and commentators hold this proposition be justified? If so, how?

II. EXISTING ACCOUNTS

Most scholars writing on coercion begin by identifying a set of core cases, which they think count uncontroversially as instances of coercion, and then proceed by offering some sort of theoretical account of them. Philosophers, for the most part, are engaged in a “definitional” exercise: they seek to establish necessary and sufficient conditions for an act of coercion, usually

a coercive speech act, and continue by first elaborating their accounts of these cases and then applying them to a set of more marginal cases.⁶ Most legal scholars, by contrast, do not attempt to supply an account of the notion of coercion itself, but instead limit their inquiry to tracing the implications of identifying an act as the product of coercion.⁷ Their interest, then, is in explaining why the fact that an agreement or an act is coerced deprives it of legal validity. They are engaged in what I call an “implicative” account.⁸ In this Article I take myself to be making a contribution to the second body of literature, with a particular focus on coerced agreements, as opposed to criminal offenses.⁹ I ask afresh whether contracts under coercion should be enforceable, and if not, why not. I argue that the usual accounts offered in the jurisprudential (i.e. legal) literature of why coercive contracts are unenforceable fail. Because the philosophical and the legal questions are not entirely distinct, however, I should make some initial remarks about the various accounts of coercion, as well as about the theory of contract invalidation each may suggest or with which each would naturally be associated. I begin by suggesting that although the common assumption that coerced contracts are invalid is correct, the reasons for this are far less obvious than both philosophical and legal scholars have previously thought. The reason for this is simple, yet it has been consistently overlooked by both philosophers and lawyers: coerced agreements are usually fully chosen by the parties, and this reflects the fact that both sides see themselves as advantaged under the agreement at the time they enter into it. For this simple, yet unnoticed reason, it turns out to be much more difficult to explain why coerced agreements are universally held to be invalid than one might suppose. My aim in this paper is to attempt to highlight the difficulties of the task, as well as to attempt to reconstruct a different approach to the problem than has generally been given.

The better account of the logic of invalidation, I suggest, can be found in an unlikely place, namely in the writings of the only philosopher who infamously actually rejected that logic altogether, Thomas Hobbes. As we will see, Hobbes has shocked many generations of scholars with his response to this problem by arguing that contracts with robbers are actually obligatory. But I argue that delving into the Hobbesian edifice a bit deeper gives us a basis for teasing a powerful argument for invalidation out of the Hobbesian contractarian framework. That Hobbes himself for the most part failed to notice this should not provide a basis for rejecting it, or ironically, for failing to credit Hobbes with its discovery.

Let us now turn to the basic accounts of coercion in the philosophical and the legal literature. The two approaches to coercion that predominate among scholars of both sorts, whether their projects are “definitional” or “implicative,” are a deontological approach on the one hand, and a consequentialist on the other. Given the divide between the definitional and implicative accounts, the further split between deontological and consequentialist approaches should in theory give us four categories: two definitional accounts, one deontological and one consequentialist, and two implicative accounts, one deontological and one consequentialist account.¹⁰

A. Deontological Accounts

To complicate matters further, deontological accounts, whether definitional or implicative, seem to come in several varieties, depending on the foundation on which the deontological claim rests. In particular, two versions of the deontological account appear to predominate: “psychological” accounts and what might call “moralized” accounts.¹¹ On the psychological version of the deontological account, coercion vitiates the agent’s ability to

choose, so that his consent to the relevant agreement, while intentionally offered, does not reflect his true will. Adherents of the psychological approach, whether philosophical or legal, often express this thought by saying that the agent's "will was overborne,"¹² that the agent's consent was offered "against his will," or that the individual was "deprived of his free will" or "mental freedom,"¹³ and thus could not truly exercise choice. Alternatively, the point is sometimes put by saying that the agreement to the contract was "involuntarily" obtained.¹⁴

Many legal scholars take the same view of duress in the criminal law, explaining the law's willingness to excuse an individual who commits a crime under duress by saying that he "did not act voluntarily" or that he was "not responsible for his actions."¹⁵ Paul Robinson, for example, writes that "[t]he excusing condition in duress is the impairment of the actor's ability to control his conduct."¹⁶ And George Fletcher says that "[e]xcuses apply on behalf of morally involuntary responses to danger; they acknowledge that when individuals *merely react* rather than choose to do wrong, they cannot fairly be held accountable."¹⁷ On the psychological version of the deontological claim coerced actions bear some resemblance to reflexive responses, in that they short-circuit the agent's rational faculties.

The moralized version of the deontological account focuses instead on the permissibility of the coercive threat and the victim's entitlement to resist the threat. The most prominent version of the moralized account is that offered by Alan Wertheimer. Wertheimer distinguishes between what he calls the "proposal prong" and the "choice prong" of coercion, namely the requirement that the threat be wrongful and the requirement that the agent's choice be impaired, respectively.¹⁸ The choice prong represents in effect that psychological element that psychological versions of the deontological account take to be impaired when an agent is subjected to coercion. While Wertheimer generally follows what he takes to be the prevailing

approach in the law, by combining both choice and proposal elements, he does not think the choice prong can carry much weight.¹⁹ Hence, Wertheimer ultimately advances a moralized account focused on the proposal prong. According to Wertheimer, the important element in coercion is the combination of the wrongfulness of the coercer's threatening proposal with the victim's entitlement to resist the threat.²⁰ On the choice side, Wertheimer calls "the principle that coercion undermines voluntariness" "uncontroversial,"²¹ but he insists on a moralized account of voluntariness itself, according to which an agent acts voluntarily just in case he acts from second-order principles to which he adheres.²² Wertheimer maintains that a threat coerces just in case it wrongfully lowers its recipient's moral baseline in such a way that it requires him to act according to second-order principles he does not truly accept, and which he is entitled to reject.²³ In this counterintuitive sense, he thinks, the coerced agent acts "involuntarily."²⁴

An alternative version of the moralized account is that offered by Robert Nozick. According to Nozick, coercion requires the coercer to have threatened a consequence that each party knows would leave the victim worse off than he would be if the coercer did not bring about that consequence.²⁵ There are further conditions, but they are not relevant for my purposes. The core of the account lies in Nozick's analysis of the notion of a "threat." Something is a threat if the consequences would make the recipient worse off than he would be in the natural or expected course of events.²⁶ Thus a central conclusion of Nozick's is that threats can be coercive but offers cannot, and in this way Nozick is able to dispose of a number of marginal cases that one might otherwise have included among the core cases of coercion.

Although the foregoing deontological accounts are not particularly focused on the implications of identifying a contract as coercive, it is generally assumed that the coercive nature of an agreement, whether merely moral or contractual, is a reason for its invalidation. But while

this point is assumed, the basis for it is never fully spelled out. For proponents of a psychological approach, the tacit assumption appears to be that coercive contracts must be invalidated because they do not reflect the voluntary agreement of the victim. The precise relationship between involuntariness and invalidation, however, is never fully articulated, and the assumption is that the invalidity of coercive contracts follows seamlessly from involuntariness, since involuntariness implies lack of consent.

In the second, more moralized account, the same conclusion would be drawn from a slightly different premise, namely that the *wrongfulness* of the coercive methods makes the victim's agreement "morally involuntary." Here the invalidation does not follow quite as automatically as it does under the psychological account. But the "moral involuntariness" of coercive contracts might easily be thought to entail the invalidity of the agreement in a way that parallels this same move in psychological accounts: involuntariness entails invalidation because it defeats true consent. And it defeats consent, such theorists think, because the consent is obtained wrongfully from the victim. There is of course something of a worry about circularity on such accounts, a worry that does not pertain to the same extent to the purely psychological accounts, as follows. The moralized accounts seem to start from the premise that the pressure applied in cases properly labeled "coercion" wrongfully deprives the victim of consent because it wrongfully lowers her baseline. But what test do such accounts use to establish whether the lowering of the baseline was in fact wrongful? This point is often left open,²⁷ but presumably proponents of a moralized baseline view would treat as wrongful any pressure that leads to morally unacceptable concessions or interactions between the one making the threat and the one receiving it. But obviously if we did not view such interactions or concessions as the product of coercion, per se, we would be unlikely to view them as morally unacceptable.

Finally, on baseline accounts, the argument for invalidation would proceed from the fact that coercion proceeds pursuant to threats that wrongfully lower the baseline welfare of the victim, and that this provides a basis for barring the coercer from claiming the gains from his conduct. Here the implication is least clear for coercive contracts, as further argument is needed to move from the central identifying features of coercion to the invalidity of coercive contracts. But such an account would not be difficult to fill in: similar to the moral account, we invalidate coercive agreements because they are premised on invalid consent.

B. Consequentialist Approaches

By contrast with the foregoing deontological approaches, the consequentialist approach to coerced contracts does not proceed by offering an account of coercion, but instead directly addresses the question of the ramifications of coercion for contract law. Consequentialists explain the invalidation of coercive contracts by the significant social harm that would come from enforcing them. They point to the obviously correct fact that if we enforce coercive contracts, all things equal, we would increase incentives to use coercion, with the result that the amount of coercion in society will also increase. Consequentialist approaches to coercion in the legal literature are economic. Richard Posner explains the unenforceability of contracts with robbers entered into under duress: “the enforcement of such offers would lower the net social product by channeling resources into the making of threats and into efforts to protect against them.”²⁸ The legal economist thus regards it as crucial that the state refuse to endorse contracts produced by illegitimate means, since otherwise it would create powerful incentives for individuals to engage in highly unproductive activity. It is a shame that this policy will sometimes work to the immediate disadvantage of a particular victim who might have otherwise

bought his way out of a life-threatening situation, but general policies cannot be made by particular cases, and the damage of such a general policy would be far too great.

Other legal economists describe coercion along similar lines. Robert Cooter and Thomas Ulen, for example, worry that enforcing coerced contracts would create an incentive for “destructive acts.”²⁹ Coercion is destructive in that it involves a “threat to destroy existing value.”³⁰ Non-coercive offers, by contrast, are threats to “block the creation of additional value.”³¹ Finally, according to Cooter and Ulen, if coerced contracts were enforced, individuals will waste resources on protection against coercion.³²

C. The Contractarian Approach

As I have already indicated, my primary focus in what follows will be the question of the implications of identifying an agreement as coercive, rather than the criteria for making that identification in the first place. What this means is that I will focus exclusively on core cases of coercion—cases that must be analyzed as instances of coercion on anyone’s theory. The accounts seeking to identify coercion are of greatest relevance for sorting through the marginal cases that philosophers like to debate, such as cases involving the possibility of “coercive offers.”

In addressing the question of the implications of coercion for the enforceability of contracts, I adopt what I call a “rational contractarian” perspective, namely a perspective from which legal policies and rules must be justified *to the individuals whose conduct falls under their sway*. The approach is contractarian insofar as it requires us to be able to posit, in at least a hypothetical sense, the agreement of such persons to the rule or policy in order for its authority over them to be justified. It is “rationalistic” insofar as the approach it will take to this question

will be a function of the perceived self-interest of the parties: rational agents would agree to a rule or policy only insofar as they regard it, broadly speaking, as in their interest. It is to be contrasted in this respect with “normative contractarian” approaches,³³ such as a Rawlsian or a Scanlonian might endorse, because agreement to a social rule or policy would not be predicated on normative factors such as the representation of human beings as “free and equal persons,”³⁴ or what human beings could not “reasonably reject.”³⁵ On a rationalistic approach, the question to ask is whether you and your assailant in the kind of coercion scenario with which I began would regard either of the above rationales—the deontological or the consequentialist—as adequate grounds for rejecting your contract. Most saliently, would you, the victim, regard either explanation as a satisfactory basis for denying your urgent plea to make your contract with your assailant enforceable? From the victim’s standpoint, the foregoing accounts will either appear flawed as accounts of coercion in their own right or will seem to provide inadequate grounds for refusing to enforce the contract.

Consider first the choice version of the deontological approach. From the victim’s point of view, the argument will seem curiously backwards. What basis is there, he might think, for accusing a person who prefers to strike a monetary deal rather than suffer immediate death of irrationality? To suggest that a person would only make such a commitment because his rational processes had been short-circuited or his will overborne seems absurd. The victim in such a case would feel that nothing could be farther from the truth, and furthermore that the person who choose not to make such a commitment rather than lose his life is the one whose rationality ought to be questioned. Indeed, any victim who was able to negotiate such a deal for herself is surely to be commended for her clear thinking rather than dismissed for lack of rationality. And this suggests that contrary to the reigning orthodoxy on this question, the victim’s action in a

case of this sort is both chosen and voluntary, despite the fact that it is not an action she would choose for its own sake.³⁶ The rational choice perspective on voluntariness, according to which an action is voluntary just in case the agent performs it for the sake of something he hopes to attain, is actually not the product of modern egoism as some philosophers would choose to portray it. It is in fact the teleological approach to intentional action that begins with Aristotle, and that authors as little interested in egoism as Elizabeth Anscombe have expounded.³⁷

Once one considers it carefully, the foregoing point is difficult to challenge. Although he wants to preserve the connection between coercion and involuntary action, Wertheimer too recognizes that coercion cannot impair voluntariness in the way that reflex reactions do, or even in the more attenuated way in which physical torture might. In standard cases of what he calls “volitional coercion,” namely coercion that operates through threats rather than physical force, the victim’s deliberative faculties are intact.³⁸ Indeed, we might say that volitional coercion *presupposes* intact rational faculties, as the threat would be wholly deprived of effectiveness if the victim lacked the capacity for rational decision making. As Thomas Schelling long ago suggested, it is in theory possible to render oneself immune to coercion simply by making oneself irrational, for example, with the aid of an irrationality pill.³⁹ The kind of coercion I have been discussing thus achieves its aim by making use of rationality rather than obviating rationality: it involves a piling on of reasons on the side of one of the courses of action available to the agent, rather than delusion, drugging, or impairment of rational faculties.⁴⁰

The confused suggestion of many philosophers and lawyers that coerced actions are involuntary is nevertheless understandable. First and foremost, they come to this conclusion via the apparently irrefutable conviction that coercive agreements must be *invalid*. Lacking any clear normative explanation for why this must be, they seek a psychological basis for their

intuitions. In this way of thinking, however, they are engaging in backward reasoning, as they are inferring the psychology from the normative conclusion it is supposed to support.

Furthermore, this mistaken move is probably further encouraged by a kind of slide from “nonvolitional” coercion, namely coercion based on physical force, to volitional coercion. For where nonvolitional coercion is concerned, it is perfectly correct that choice, decision making, and reflection have no role to play in producing the actions of the agent. In such cases, the agent becomes a mere instrument of someone else’s will. Yet volitional coercion is significantly different, insofar as an agent still selects a course of action because he sees it as in some way better for himself.⁴¹ Rational choice theorists, then, seem correct in their assumption that bad choices are not to be equated with non-choices, and a powerful incentive to do one thing rather than another does not make the agent’s subsequent actions in any meaningful sense involuntary.

For the foregoing reasons, the moralized versions of the deontological approach seem at first blush more promising. I will consider Wertheimer’s and Nozick’s accounts together, since they have a common structure. In both accounts, the assailant has impermissibly lowered the baseline of the victim by putting her into a situation in which she is forced to make a highly undesirable choice: a choice to avoid a situation that is below a certain moral baseline or a choice to avoid a situation that diverges from the “natural or expected course of events,” which sets a non-moral baseline.⁴² Both Wertheimer and Nozick recognize moral and non-moral baselines. Moral baselines enter the discussion to address the problematic case of the slave who is threatened with a beating unless he does something for his owner. The complexity of the case stems from the fact that the slave can expect a beating tomorrow no matter what, so there would be no divergence from the natural and expected course of events (the non-moral baseline). Using a non-moral baseline in this case would suggest that it is not an instance of coercion, but our

sense that the threat is highly coercive suggests that a moral baseline might be more appropriate.⁴³ Nozick and Wertheimer diverge, however, in how they think we should determine which baseline is appropriate to use in which circumstances. Nozick believes that the appropriate baseline is the baseline that the potential victim himself would *prefer* were used in assessing his subjection to coercion.⁴⁴ Wertheimer believes that the appropriate baseline is determined by *context* and our *overall moral theory*.⁴⁵

At this point, both views fall into the problem that I have been stressing. Regarding Nozick, if the victim's preferences matter in choosing the appropriate baseline, the victim's preferences should also matter in the selection of the correct moral and legal response to the coerced contract. And in Wertheimer's case, if our overall moral theory is what determines the choice of baselines, a moral theory that required that a proposed rule be justified to a victim might have the result of suspending a duress defense in contract law rather than supporting it. Consequently, considerations that go into identifying the proper baseline seem ultimately to suggest that identifying a case as an instance of coercion does not automatically generate the implications of such an identification.

Thus Wertheimer and Nozick, whatever their differences, would presumably both want to reject the suggestion that the victim is made better off by entering into a contract with the assailant. The sense of "better off" in this case is an artificial product of the assailant's wrongful manipulation of the victim's options, and this gives us a basis for depriving the assailant of the gains from his coercive tactics. Since coercion impermissibly lowers the victim's baseline by restricting her options in a way that she strongly disprefers, we cannot think of the agreement as consented to, even hypothetically, by the victim.

But if we take up the victim's standpoint once more, we can see that the foregoing moral

versions of the deontological argument will be no more compelling than the psychological version. “True,” the victim will say, “I did not want to be placed in this position, and would of course rather not have had to face the choice between my money and my life. And it is also true that the means my assailant used to place me in this position was wholly illegitimate: using force to limit my options and thereby improve his own position is the clearest example of an improper means for gaining an advantage over another imaginable. But none of that is particularly relevant to me now,” she might say, “for focusing on the way I came to be in this position does not now benefit me one bit. Those features of my situation are in effect “sunk costs.” My only interest now is how I can save my life, and the best way to do that is to enter into a binding agreement that commits me to return with money the next day. So although I agree that it is extremely unfair of my assailant to have put me in a position where I had to resort to such an agreement, it seems doubly unfair to now deprive me of the only means I have to save my life under such circumstances.” The victim has a point. If the law focuses on the wrongfulness of the assailant’s conduct to the exclusion of her welfare, it will be in effect conspiring to lower the victim’s baseline even further than the assailant would have lowered it if left to his own devices. The moralized version of the deontological approach thus seems no more promising from the rational contractarian perspective than the psychological version was, as it leaves the victim’s plea entirely unanswered.

It might be tempting to suppose that the consequentialist approach to this problem would best capture the rational choice perspective, since deterrence is an important concern for rational agents in the context of social agreement. But a moment’s reflection will make clear that the appeal to deterrence also fails to capture the viewpoint of rational agents who are to be subject to the rules governing coercive contracts, and so fails to respect the most basic principle of

contractarian approaches, namely that we can represent the resulting social arrangement as *contractual* because we can infer agreement from the fact of mutual benefit. Thus, for example, the economic version of the consequentialist solution says that we must reject coerced contracts because enforcing them would produce overall lower social utility than rejecting them, *despite* the fact that from the standpoint of the individual parties, enforcing the contract may be Pareto optimal. More generally, a consequentialist might say that neither party would want to live in a world in which anyone can make extensive use of their superior threat advantage and receive legal blessing for doing so. The lives of all of us would be less secure if such agreements were rendered enforceable. Consequentialists therefore assert that the plight of particular victims must therefore yield to the sound policy of deterrence. It is a shame that this policy will have its casualties, but there is no other reasonable way to guard against the specter of massive increases in coercive demands that failing to adopt this policy would produce.

Once again, from a rational contractarian perspective, there is a powerful basis for rejecting this rationale as a reason for refusing to enforce coercive contracts. Addressing the situation from the victim's perspective, it will be hard to see enhancements to overall social utility as an adequate basis for justifying the adoption of a policy that will harm *her*. She will rightly see herself as a sacrificial lamb: "It is bad enough that I was put in this awful position by my assailant," she might say, "but now I am being condemned to death for the sake of a social policy that *will benefit everyone else but me!* Indeed, even my coercer will benefit from this policy, since he will live to tell the tale and so to enjoy the future benefits of this deterrence, but I, whose life must now end, will not." And so once again, from a rational contractarian perspective there is reason to conclude that the deterrence account is seriously lacking in necessary complexity, because it fails to supply a justification that the victim in this scenario

would have reason to see as compelling.

This line of attack raises a fundamental question about what would count as an adequate justification for the policy of non-enforcement of coercive contracts. I have suggested that such a policy cannot be justified *to* the victim who must suffer its consequences, and thus fails the rational contractarian requirement that the parties who must live under the rule see something for themselves in it if its authority over them is to be legitimate. To summarize the difficulty, the problem in this case is that the diffuse personal benefit each may receive from an enhancement to overall social utility may *not* strike rational agents as adequate benefit to counterbalance whatever personal loss of utility the rule may entail.

In this context, the question from a contractarian perspective is *not* whether the coercer has impermissibly lowered the victim's baseline below what it would have been without his intervention, or whether social utility would, in the aggregate (or even on average) be enhanced by refusing to enforce coerced contracts. The question is rather *whether individuals selecting the terms of their own legal system would choose to include a duress defense that invalidated coercive contracts or whether they would choose instead to eliminate that defense and allow such contracts to be enforced*. The relevant question, in short, is precisely the one raised, but not answered, in the imaginary conversation between the victim and the robber: if the parties were to sit down together long in advance of any need to enter into a coercive agreement, would they choose to retain the duress defense or would they choose to eliminate it?

Note that economists sometimes pay lip service to the contractarian approach by suggesting that it is *evidence* for the truth of the economic conclusion. Posner, in his remarks about coerced contracts quoted above, continues by saying: "We know that this class of "contracts" is non-optimal because *ex ante*—that is, before the threat is made—if you asked the

Bs of this world whether they would consider themselves better off if extortion flourished, they would say no.”⁴⁶ But it is far from clear that Posner is correct in this regard. In effect, despite the respectful remarks made about the contractarian thought experiment, legal economists do not take seriously the individualistic perspective that the contractarian approach in fact requires, and so in the end would not be prepared to accept the upshot of any actual or even putative agreement among rational agents.⁴⁷

How would we go about determining what the parties to a potential ex ante “coercion” agreement would decide? The question is whether the parties would consider themselves better off living with the added deterrent benefits that making coercive contracts unenforceable would provide, or whether they would prefer to protect their ability to buy their way out of coercive situations in case they found themselves subject to such coercion. From an ex ante perspective, the answer is far from clear. Given that the parties are aware prior to the moment of contracting that they may need to save their lives by making an enforceable promise to a robber, and given the intensity of their desire to make binding promises under such circumstances (which they can anticipate), the balance of personal utilities *could* weigh in favor of allowing such contracts, and so the parties might select a regime that enforced coercive contracts, even in the absence of the immediate need to enter one. The rational course of action would depend on the increased likelihood of coercion, the effectiveness of the promises one could make under such circumstances, and so on. For the rational choice theorist, the answer to the question of coercive contracts thus appears to be a *highly* contingent one.⁴⁸

To make the suggestion that the parties *might* opt for protecting their ability to buy their way out of coercive situations more plausible, consider the conditions that obtain in some countries today of frequent kidnappings for ransom. In such countries, kidnapping is usually

economically motivated and victims are nearly always released if the ransom money is provided. Because law enforcement is weak and the chances of prosecution low, the victims are virtually certain to be killed if the ransom is not provided. Clearly there are significant utilitarian advantages to the position that kidnappers should not be rewarded with ransom money. If such a principle were consistently adopted, there would be fewer kidnappings in that country. It is equally clear, however, that once one is the victim of such a kidnapping, one would desperately hope that one's country did *not* follow a no-cooperation-with-kidnappers policy. What this kind of case reminds us, then, is that there is a potentially significant gap that obtains between *personal* utility and *social* utility calculations. There appears to be no particular reason to assume that the maximizing solution from the standpoint of social utility will be the same solution that rational agents seeking to maximize their own utility would adopt, if unanimous or close to unanimous agreement is required. For this reason, the utilitarian and the contractarian solutions to social problems will often diverge.⁴⁹

In this case, the utilitarian solution would likely support a no-cooperation-with-kidnappers policy.⁵⁰ But what would the rational contractarian solution suggest? The relevant question, once again, is what a rational agent would say about the decision to adopt the no-cooperation policy if asked *prior* to being kidnapped, knowing there is a certain likelihood of finding himself a kidnapping victim with no other way. On the one hand he must consider the effects on future kidnapping victims of entering into bargains with kidnappers, as well as contemplate the likelihood that he would be among them. With only these two factors to consider, it would not be surprising if he reached the same conclusion as the utilitarian social planner, given that the loss of deterrence might be severe, while the chances of being kidnapped for a single individual might remain low. But when, on the other hand, he considers the gravity

of the potential evil, and considers the importance of preserving his ability to buy his way out of a kidnapping situation he may reach a very different conclusion. In light of these considerations, it is far from obvious that from an ex ante perspective one would side with the more abstract social policy considerations over one's personal preferences as one imagines them to be ex post, even though the individual is also the beneficiary of the social policy.

A further argument, however, may transform the nature of the perceived benefit and provide a definitive reason for the rational agent to favor protecting his ability to save himself from dire, but unlikely situations, and hence to reject the consequentialist approach. In many situations it seems possible to achieve the deterrent effect one seeks from a policy of non-cooperation or non-enforcement by simply increasing the penalties for the initial coercion itself. Assuming that penalties can always be further increased, there is no reason to think that the deterrent benefit on which the utilitarian argument depends *must* come from invalidating the contract (as though invalidation were the only available way to create disincentives). In cases in which independent punitive measures are possible, the individual utilities from allowing coercive contracts can be assessed quite separately from the deterrent benefits from invalidating coercive contracts. There would then be little individual benefit to banning coercive contracts or refusing to cooperate with kidnappers, and, on the other side, significant individual security from retaining the ability to enter into such contracts or engage in such negotiations.

For the foregoing reasons, the deontological and the consequentialist arguments against enforcing coercive contracts have significant weaknesses, particularly from a rational contractarian perspective. While the objections to these approaches may be compelling, developing an affirmative alternative approach to the problem is quite another matter. I will be aided in this task by considering the remarks of a thinker who adopts the rational choice

perspective, namely those of Thomas Hobbes. It is perhaps not surprising that Hobbes is also the only philosopher to have written on this subject who thinks that, at least in some circumstances, contracts under coercion should actually be enforceable. I will use Hobbes' remarks on this topic to consider whether the contractarian approach is inevitably committed to this counter-intuitive position, or whether the more common sense conclusion that coercive contracts are unenforceable might yet be justified in rational choice terms.

III. HOBBS AND THE ROBBER

A. Coercive Contracts in Nature

Against what is surely the nearly universal position on coerced contracts, Thomas Hobbes stridently defends the claim that agreements entered into under duress or coercion are binding. His first defense of this position appears somewhat confusedly in *De Cive*, where he writes as follows:

It is a usual question, whether compacts extorted from us through fear, do oblige or not. For example, if, to redeem my life from the power of a robber, I promise to pay him 100*l.* the next day, and that I will do no act whereby to apprehend and bring him to justice: whether I am tied to keep promise or not. But though such a promise must sometimes be judged to be of no effect yet it is not to be accounted so because it proceedeth from fear. For then it would follow, that those promises which reduced men to a civil life, and by which laws were made, might likewise be of none effect...⁵¹

Hobbes' position comes substantially into focus when he takes up this topic in XIV of

Leviathan. He first addresses this question in the context of agreements entered into in a state of nature to free oneself from the grip of an enemy:

Covenants entered into by fear, in the condition of mere nature, are obligatory.

For example, if I covenant to pay a ransom, or service for my life, to an enemy, I am bound by it. For it is a contract wherein one receiveth the benefit of life; the other is to receive money, or service, for it; and consequently, where no other law (as in the condition of mere nature) forbiddeth the performance, the covenant is valid.⁵²

That is, Hobbes' reasoning is that the enemy and his victim exchange items of value, and each receives a benefit from the other in so doing. For Hobbes, the benefit received makes the commitment fully voluntary, as we have already seen. There is thus no psychological basis for invalidating the contract. He explains:

The cause of fear which maketh such a covenant invalid must be always something arising after the covenant made (as some new fact or other sign of the will not to perform), else it cannot make the covenant void. For that which could not hinder a man from promising, ought not to be admitted as a hindrance of performing.⁵³

It is not the fear itself that invalidates the agreement, but the cause of the fear, *as long as that cause has to do with new circumstances not apparent at the time the contract was made*.

Significantly changed circumstances or acts of God, as they are typically called, might make a contract invalid, but only if unavailable to the parties at the time of contracting. Hobbes' basic thought is that circumstances available to the parties at the time of the agreement do not impair the parties' obligations if they did not impair the parties' ability to exchange promises in the first

place.⁵⁴ And this implies that promises made as part of an exchange always obligate the promisor, as long as they are voluntarily made, in Hobbes' sense.⁵⁵

B. Coercive Contracts in Civil Society

Hobbes takes the same position with regard to the *De Cive* example, namely the case of agreements made with robbers in *civil* society. And as Hobbes also points out in the *De Cive* discussion of this point, the sovereign may have good reasons for forbidding the robber's conduct separately, which would provide a conventional limitation on one's natural obligations:

[I]n commonwealths, if I be forced to redeem myself from a thief by promising him money, I am bound to pay it, till the civil law discharge me. For whatsoever I may lawfully do without obligation, the same I may lawfully covenant to do through fear; and what I lawfully covenant, I cannot lawfully break.⁵⁶

Hobbes thus thinks an agreement entered into with a robber in civil society, made to save one's life is binding, as long as the law does not forbid such contracts. But since he also thinks the civil law should be a translation of the natural law into civil form, we can suppose that he probably thinks the law *should* enforce such contracts. At any rate, no civil government has a duty to invalidate contracts of this sort for moral reasons, Hobbes firmly believes.

Hobbes' account of why and when coercive contracts are binding in civil society is largely the same as the one he offers for the state of nature. In speaking of "whatsoever I may lawfully do without obligation," Hobbes is suggesting that a person can bind himself to anything, as long as he is free to do or not to do it. And this suggests that agreements entered into out of fear are binding, as long as the promised actions are not otherwise the subject of obligations. Thus acts that are already obligatory, one cannot now make *more* obligatory by promising once

more to perform them, and acts that are the subject of prohibitions, such as illegal acts, one cannot bind oneself to perform. Once again, the relevant logic is that what does not impair the promise should not impair the contract. In this case, the agent is free—not impeded—from giving a robber money if he wishes. He is not already obligated to do so, nor is he forbidden from doing so. And what he was free to promise he is also free to impose on himself as an obligation.

Lest one suppose that Hobbes is not serious about these remarks about the robber, the same idea is recapitulated in his discussion of sovereignty by acquisition, according to which “the sovereign power is acquired by force.”⁵⁷ The sovereign power in such cases is created when men “for fear of death or bonds do authorize all the actions of that man or assembly that hath their lives and liberty in his power.”⁵⁸ Hobbes goes on to explain in this passage that the only difference between sovereignty by institution and sovereignty by acquisition is that in the former case men choose their sovereign “for fear of one another,” whereas in the latter case, it is for fear “of him whom they institute.”⁵⁹ As Hobbes points out, the two sorts of sovereignty are ultimately not too different, since in both cases “they do it for fear,” which he thinks is worth noting as against people who erroneously “hold all such covenants as proceed from fear of death or violence void.”⁶⁰

Further, Hobbes argues, if fear did indeed render all agreements void, “no man in any kind of commonwealth could be obliged to obedience,”⁶¹ i.e. no man would be obligated to obey the law. For example, Hobbes says elsewhere, a man might pay a debt only because he fears imprisonment.⁶² But he nevertheless pays the debt freely, and he was at liberty not to pay it. Hobbes is suggesting, *reductio*-style, that coercion could not negate freedom because we would then have the absurdity that we could no longer see anyone as obeying the law of his own free

will. Coercion, like law more generally, merely changes the payoffs for rational agents. It does not deprive those agents of their ability to choose.

Hobbes' argument on this topic is striking, not just because of the counterintuitive nature of its conclusion, but also because it seems to be at odds with one of the central premises of his political philosophy, namely the idea that in a state of nature, there are no obligations of any kind. As Hobbes famously says in Chapter XIII of *Leviathan* in describing the state of nature: "The notions of right and wrong, justice and injustice have there no place. Where there is no common power, there is no law; where no law, no injustice."⁶³ It follows from the Chapter XIII remarks that no contract in nature can be binding, as there is no power to enforce any agreement. How, then, does Hobbes reconcile his claim that contracts with robbers are binding in a state of nature with his equally clear statements to the effect that contracts in a state of nature are not in general binding?

Moreover, if the general appeal to the invalidity of contracts were not compelling enough, Hobbes makes a more specific point that should compound any confusion about the case of the robber. He says, "he which performeth first does but betray himself to his enemy, contrary to the right (he can never abandon) of defending his life and means of living."⁶⁴ From this one would have expected the following line of reasoning: since the robber, as first performer, merely "betrays himself" to his victim, and cannot therefore expect promises made to him to be kept, his victim has no obligation to keep the promise, since it would not have been rational for the robber to rely on him in the first place. This would seem to be the obvious conclusion from a rational choice perspective, and it would have perhaps been a sensible thing for Hobbes to argue. But curiously, it is not the view he took. Instead, he appears to be saying that although contracts in nature are generally not binding, contracts with robbers *are*. But how could any sane person

defend such a claim? Are we sure Hobbes really believed it?

C. Contracts and Covenants

The explanation may lie in a distinction Hobbes makes between two kinds of contracts, which I call “first-performer” and “future performance” contracts,⁶⁵ as well as in the argument that Hobbes uses against the infamous fool, whose challenge to the third law of nature provides a pivotal insight into Hobbes’ political philosophy more generally. Considered carefully, it appears that Hobbes is not in fact committed to the blanket suggestion that contracts in a state of nature are invalid. Hobbes actually thinks that some contracts are rational to keep, even in the absence of enforcement or moral obligations. These contracts he calls “covenants,” in order to distinguish them from ordinary contracts, which he calls “contracts.” The former is a contract in which one party performs at the time the contract is made and the other party commits to performing some time in the future,⁶⁶ whereas the latter is one in which both parties commit to performing at some time in the future.⁶⁷ It will therefore turn out that the contract with the robber is a species of first-performer contract, or “covenant,” and as such provides an exception to the general rule that contracts in nature are not binding.

Many commentators have overlooked the distinction Hobbes was trying to draw between covenants and contracts. Among the commentators that have focused on the distinction, the usual view is that a covenant is just an ordinary exchange of promises for future performance of some sort. Howard Warrender, for example, says that a covenant is when “one *or both* parties are to perform in the future and are to be trusted in the meantime.”⁶⁸ Similarly, Gregory Kavka says that “Contracts in which one or more parties are called on to perform their parts at some time after the contract is made are *covenants*.”⁶⁹ And David Gauthier says that “A covenant is

that species of contract in which at least one of the parties 'is to perform in time to come.'"⁷⁰

There are good reasons, however, not to read the notion of a covenant in this open-ended way.

First, in the passage where Hobbes introduces the notion of covenant, he clearly has in mind a case in which one party performs now in exchange for a promise on the other party's part to perform in the future. He writes: "Again, one of the contractors may deliver the thing contracted for on his part, and leave the other to perform his part at some determinate time after (and in the meantime be trusted); *and then* the contract on his part is called PACT or COVENANT."⁷¹ And second, the foregoing interpretation leaves commentators with no way of explaining the distinction between covenants and ordinary contracts, despite the fact that they notice Hobbes drawing such a distinction. They cannot, for example, say that covenants are that species of contract in which at least *one* party must perform in the future, leaving the larger class of contracts to include agreements in which both parties perform at the time the contract is made. For Hobbes says a contract requires a "mutual transferring of right," and for this reason, he does not think of barter as a contractual exchange.⁷² Hobbes gives us no other hint about how to draw the distinction between covenants and ordinary contracts, and it is hard to see any reasonable alternative.

Now there is a distinction in modern law that roughly corresponds to the distinction between future-performance and first-performer contracts, and this is the distinction between so-called "bilateral" and "unilateral" contracts. Bilateral contracts are contracts in which parties exchange promises, and both performances are to occur in the future.⁷³ Unilateral contracts are contracts in which one party exchanges a current performance for another party's promise.⁷⁴ The distinction between bilateral and unilateral contracts is very close to the distinction Hobbes is trying to draw between contracts and covenants, but does not exactly track it, since what Hobbes

is concerned to condemn is *being a first performer*, in any type of contract in which it might be required.⁷⁵ And this suggests that it would be rational to act on a bilateral contract if one's performance were required second, rather than first, as long as the other party had already performed. On the other side, Hobbes would presumably be prepared to reject the rationality of entering into a unilateral contract, if the terms of the contract required one to be the first performer. We cannot, then, simplify and say that Hobbes regards bilateral contracts as void but unilateral contracts as valid in a state of nature, since the distinction he traces is that between being a first performer versus being a second performer, a distinction that cuts at right angles the distinction between bilateral and unilateral contracts. What, then, *are* the implications of Hobbes' view of being a first performer for the validity of contracts in nature?

First consider bilateral contracts. If it is irrational to be a first performer, then it would be rational to reject any exchange of promises in which both parties are to perform in the future, as it would be irrational to commit to first performance, and naïve to commit to second performance (since the first performer will almost certainly never perform). So a mere exchange of promises cannot be a valid contract in nature. Thus the condemnation Hobbes offers of the rationality of being a first performer will invalidate the rationality of bilateral contracts in nature across the board. The fact that Hobbes allows that being a second performer can be rational, then, does not help to establish the rationality of bilateral contracts.

What about unilateral contracts? Arguably, on Hobbes' assumptions, it is *rational* to offer to exchange a promise for another's performance. According to black letter law, offers for unilateral contracts are deemed accepted only upon performance.⁷⁶ Thus if A offers to pay B \$500 if B climbs to the top of a flagpole, B is not deemed to have accepted until he reaches the top, and correspondingly A is not bound to pay until B's performance is complete. Thus A's

obligation would commence only once he received the benefit he was trying to achieve. And this makes second performance very different in a unilateral contract than in a bilateral contract, given that the second performer can promise to perform in such a contract without risk. From this it would appear to follow that it is rational to be the offeror or promisor in a unilateral contract by Hobbes' lights, but not to accept another's offer for a unilateral contract by performing first.

Hobbes' defense of the rationality of keeping first-performer contracts appears in what might seem an unlikely place, namely in the response he offers to the fool's challenge to the third law of nature. The third law of nature, to recall, is presented as a blanket injunction to keep covenants:

From that law of nature by which we are obliged to transfer to another such rights as, being retained, hinder the peace of mankind, there followeth a third, which is this that men perform their covenants made, without which covenants are in vain, and but empty words, and the right of all men to all things remaining, we are still in the condition of war.⁷⁷

Notice that the third law of nature, as stated, applies to *all* contracts, and not only to first-performer contracts. My suggestion in the preceding pages, however, has been that Hobbes intends it to apply only to first-performer contracts. While exploring Hobbes' remarks in these well-known passages may seem something of a digression, understanding his argument on this point turns out to be crucial for analyzing his defense of contracts with robbers, and in particular for understanding how such contracts are to be distinguished from the mass of ordinary contracts in nature.

D. Hobbes' Answer to the Fool

The Fool argues that there can be no general rule of reason that advises men to keep their covenants, since it can only be rational to keep covenants when it would leave one better off to do so.⁷⁸ On this view, the rationality of keeping covenants is entirely an empirical matter: there simply is no *a priori* way to rule out the possibility of cases in which it is rational to break covenants. Now Hobbes allows that the Fool is generally correct about contracts in nature. But he insists that the Fool's position does not apply to second performance on a first-performer contract:

For the question is not of promises mutual where there is no security of performance on either side (as when there is no civil power erected over the parties promising), for such promises are no covenants, but either where one of the parties has performed already, or where there is a power to make him perform, there is the question whether it be against reason, that is, against the benefit of the other to perform or not. And I say it is not against reason.⁷⁹

Consistent with the argument about bilateral and unilateral contracts I considered earlier, in this passage Hobbes is associating cases in which one party has already performed with cases in which there is a common power over the parties to make them perform. That is, first-performer contracts are treated as on a par with contracts entered into in civil society. And such contracts, Hobbes explains, are not subject to the general invalidity of contracts in nature, and hence are exempt from the Fool's challenge.

Hobbes gives a rather elusive explanation for this surprising claim. First, he suggests that antisocial acts that normally tend to one's own destruction cannot be rationally vindicated just because they turn out to be advantageous on particular occasions. The fact that one can get away

with violating a contract every once in a while does not mean that habitual contract breaking is a rational thing to do. Hobbes then continues in this vein but adds a further argument:

He, therefore, that breaketh his covenant, and consequently declareth that he thinks he may with reason do so, cannot be received into any society that unite themselves for peace and defence but by the error of them that receive him; nor when he is received, be retained in it without seeing the danger of their error; which errors a man cannot reasonably reckon upon as the means of his security . . . and so as all men that contribute not to his destruction forbear him only out of ignorance of what is good for themselves.⁸⁰

What is Hobbes' argument here? One way to understand it is to see it as a form of a *reductio*. Suppose it *were* rational to be a covenant-breaker in an expected benefit sense, because, let us suppose, the chances of detection were sufficiently low, and the benefits of being a free rider sufficiently high, that it seemed rational to issue a deceitful promise. Assume that others have the same access to the rationality of one's conduct that one does oneself—in other words, that if a certain course of action is rational, and is known to the actor to be so, then others will understand it to be rational as well. This assumption is an off-shoot of what game theorists call “common knowledge of rationality.”⁸¹ It follows that if it were rational to covenant to perform and nevertheless defect, others would know it was rational to do so. The result would be that they would not expose themselves by covenanting with you, and you would end up excluded from cooperative arrangements. Against the background of the common knowledge assumption, the first performer would not have performed in a unilateral contract unless it were rational for the second performer to keep his promise and perform in turn. If the first performer *has* performed, then, it follows that he has taken the second performer to be a rational agent, and

knowing himself to be such, the second performer has every reason to perform.

If I am correct in my reading of Hobbes' response to the Fool, Hobbes is here endorsing the conditional strategy known to game theorists as "tit-for-tat." Tit-for-tat is the strategy of responding to one's fellow player by always repeating what the other player does.⁸² An objection will arise at this point, however, that tit-for-tat is unnecessary, and indeed irrational, in a situation in which common knowledge of rationality obtains. For if one knows with certainty what the other player believes to be rational and one also knows that action to be rational, then one can perfectly predict the other person's conduct. In this case, cooperation will often be rational, at least in situations in which there are mutual gains from cooperation. But tit-for-tat would be an excellent strategy for a situation in which there is less than perfect certainty about the existence of common knowledge, and hence less than perfect certainty about whether one is dealing with a fellow cooperator. Arguably this is precisely the situation that obtains if parties in a state of nature are mostly, but somewhat inconsistently rational. If one is uncertain whether the other party is rational, it makes sense to refrain from engaging him until one has evidence about the state of his rationality. The problem with a bare exchange of promises, then, is that one cannot be certain at the time one must render one's commitment that the other party is rational and is prepared to behave cooperatively. One would thus be obliged to commit oneself to performance in the absence of knowing the other party's (true) intentions. This is true even if one is contemplating being a second performer in a future performance contract. For in such a case, *at the time one's commitment is made*, one does not know whether the other party is a cooperator. It is true that one can always make one's commitment conditional on the other party's performance, and thus establish that one's obligation to perform on the contract will cease if the first performer fails to act. Yet a conditional agreement is either a form of first

performer contract (“I will pay you \$100 if you climb to the top of the flagpole”) or a promise to do something in exchange for the other’s promise. And Hobbes’ point is that in the absence of knowledge of the other person’s rationality, it cannot be rational to make a commitment based on the receipt of a mere promise on the other party’s part.

Clearly being the promise, or second performer, in a unilateral contract stands on a different footing. In this case, one can be assured that the first performer is rational *before* one is obliged to make a commitment. Or rather, one’s commitment is conditional on the fact of his *actual* cooperation, rather than a mere promise of cooperation. And thus the promiser in a unilateral contract is exposed to no risk by making a promise that is conditional on the fact of the other party’s performance. This, then, is the basis for distinguishing first performer from second performer contracts in nature, and hence for Hobbes’ willingness to exempt second performance on a unilateral contract from the general invalidity of contracts in nature.

The argument Hobbes makes to the fool can essentially be applied directly to the robber case to explain his views on this example: if it was rational for you to renege on the promise to bring the money, the robber would know it was rational for you to do so, and would not have trusted you by releasing you. So the robber’s willingness to release you from his grip, in exchange for your promise to return with the money, suggests that it *is* rational for you to return with the money, in accordance with a plan that leaves both of you better off than you would otherwise be.⁸³ And notice, by the way, that the contract with the robber is a unilateral contract: the robber trades a present performance on his part—releasing his grip on you—for your promise to return in the future with the money. Since you are not the first performer, it is rational for you to keep your end of the bargain, given that there is no danger that you will not receive the benefit under it that you were due.

What about the arguments from baselines I considered in connection with Wertheimer's and Nozick's accounts? A baseline theorist would want to say that the victim's promise is invalid because it was the product of an artificially lowered baseline that the coercer himself produced. But for Hobbes, it is the archetype of the rational exchange in a state of nature that a person is able to reap the fruits of his domination over another when both begin in a position of equality. Whatever concessions the threatening agent can extract from the other are fair game, and a stronger or more able agent should not be deprived of the benefits of his superior abilities merely because they work to the disadvantage of the weaker party. For Hobbes, then, the coercive, unilateral contract reflects the returns to the performer of his superior position in the state of nature. Since coercion supplies a rightful basis for gaining concessions from another in a state of nature, and since it is often rational to surrender power in the face of coercive tactics in nature, the agreement with the latter in effect supplies the paradigm of the binding agreement in nature. If the other voluntarily agrees to some mutually beneficial course of action in order to mitigate the effects of that exercise of coercive force, the first performer is entitled to demand the other's fidelity to that agreement, since he, the first performer, has cashed in his superior position to secure it.

E. Exploring the Hobbesian Framework

All that has just been said about Hobbes and coercive contracts will no doubt only serve to hasten the speed with which most readers will be inclined to dismiss a rational choice approach to such a topic, at least if such approaches require us to adopt a Hobbesian stance. For modern readers—like Hobbes' contemporaries—will reject a political philosophy that legitimizes coercers taking advantage of their superior threat advantage. Indeed, it is usually

regarded as one of the fundamental advances of modern contract doctrine that it recognizes inequality in bargaining position as a basis for voiding a contract. Thus far from allowing individuals to claim a rightful entitlement to the fruits of a superior bargaining position, modern legal scholars, as well as contemporary philosophers, will see the argument from domination as an additional reason to treat such contracts as void. Furthermore, they will assert, Hobbes' commitment to the foregoing argument supplies all the more reason to question his political philosophy across the board. Thus Hobbes' account of contracts under coercion seems a microcosm through which to question the entire Hobbesian edifice.

Still, let us consider carefully whether this response is warranted. In particular, two features of Hobbes' analysis give us reason to suspend judgment and investigate further. First, remember the plight of the victim and her ardent wish to be able to suspend the duress defense in order to enter into a valid "coercive" contract. In defending coercive contracts, Hobbes is not only granting the coercer permission to take advantage of his superior bargaining power. He is also accommodating the victim's preferences in a way that needs to be taken seriously from a rational choice perspective. Contemplating this aspect of such situations, even Wertheimer is forced to admit:

[I]f B would prefer that A make a freedom-enhancing proposal, given a plausible understanding of B's present situation, and if, at least prospectively, B would prefer to be bound by his agreement, it is arguable that there should be at least a reasonably strong presumption that the agreement is not coerced and ought to be enforced.⁸⁴

Of course Wertheimer puts the point by saying that in such cases there should be a presumption that the agreement was not coerced. Of course we might as well say that the

agreement was valid, though coercive, rather than use its validity to deny that agreements with these features are coercive. A more serious problem with Wertheimer's admission here, however, is that he fails to notice that B's preference to be bound by the agreement is not an occasional, outlying feature of coercive agreements. It is the hallmark of such agreements, at least those that result from so-called volitional coercion. The general structure I have identified in the gunman scenario is a fully generalizable aspect of coercive contracts: ex post the victim will *always* prefer such contracts to be enforceable. Ex ante he may or may not, and thus Wertheimer's gloss to that effect does not strengthen his account. But if even a deontologically oriented theorist like Wertheimer can feel the force of the victim's ex post wishes and their relevance for ex ante decision making, we have reason to consider approaches that give such preferences their due. Of course this point does not, by itself, provide anything like an all-things-considered argument *in favor* of enforcing coercive contracts. It is only one consideration that might be raised in support of such an argument. What we should hope to find in the end is a way of respecting the victim's preferences, at the same time that we are able to bring Hobbesian legal philosophy into synch with contemporary intuitions of justice, and thus find a way of justifying, rather than embracing, the invalidity of coercive contracts in the face of the rational choice arguments.

Second, recall that Hobbes' views on this question are meant to apply *in the absence of any law to the contrary*. Thus the entitlement created by coercive power is a natural entitlement only, one that persists in civil society only in the absence of sovereign interdiction. What, exactly, a Hobbesian has to say about coercive contracts in civil society, and in particular what the possible justifications for forbidding coercive contracts might be, requires more careful exploration. Hobbes left us something of a black box in that regard, and we need to extract the

Hobbesian position on coercion in civil society from other aspects of his political and legal philosophy. It is to this task that I now turn.

IV. TOWARDS A CONTRACTARIAN APPROACH TO COERCIVE CONTRACTS

In the imaginary discussion with the robber, I argued that any effort to make a coercive agreement enforceable by contractually suspending the duress defense is subject to a fatal flaw: since the agreement to suspend the ban on duress would itself be entered into under duress, there would be no basis for exempting it from the general rule that voids contracts under coercion. I started to consider whether matters might be different if the parties had an opportunity to address such an agreement prior to the coercive interaction. Let us now return to that question. Would parties who are presently unaware of whether they will be victims of coercion, or whether they will want to exercise coercive power over others, decide to make coercive contracts enforceable? Or would they choose to ban them, as current policy suggests we do?

Previously I asked this question in the isolated two-party case, and we imagined that you and the robber would answer it by prior agreement, without the benefit of knowledge as to what your future positions might be. We might now imagine that we have a group of individuals choosing rules for the establishment of a legal system, faced with a choice among different possible regimes: would rational agents select a regime in which coercion is generally permitted, where gunmen are allowed to extract benefits from others by force, where coercive agreements are upheld, coercive transfers of property respected, coercively sought sexual favors permitted, and so on? Or would they prefer a regime in which coercion is rejected in all of its uses and manifestations—where the state sought to punish those who attempted to gain from the private threat of force, and where it deprived those who did resort to coercion of the benefit of those

efforts, despite the fact that they were merely exploiting natural advantages?

It is not hard to see that even on Hobbesian premises, the regime that permitted coercion would be ruled out, as it would constitute a near-certain reversion to a state of war. And this gives us a basis for thinking that Hobbes can be rescued from the perils of moral perfidy. He can simply say that although agreements with robbers are binding under natural principles, the sovereign has good reason to forbid the use of coercion in civil society. This would square with Hobbes' remark that where "the thing promised is contrary to the laws,"⁸⁵ for example where "force is prohibited," the actions used to induce agreement are impermissible, and so the agreement would be void.⁸⁶ Will this quick and dirty solution prove an adequate route back to common sense for a Hobbesian? In my view it will not, at least not without further elucidation.

Hobbes' implicit argument here appears to be that if the sovereign has good reason to illegalize coercion, it follows that he has good reason to bar coercive contracts. But the suggestion depends upon Hobbes' conflation of two quite different questions that must, at least in the first instance, be kept apart: the question of the illegality of the *content* of a promise, versus the question of the illegality of the *actions used to induce* the promise.⁸⁷ In principle, we might have varying responses to these different types or sources of illegality. One can imagine a situation in which the conditions creating the coercive effect are not illegal (I threaten to break off my friendship with you unless you promise to do such-and-such), but the content of the promise *is* (you must promise to do a drug run for me). Alternatively, the conditions creating the coercion might be illegal (I threaten to kill you unless you do such-and-such), but the demanded action is not (you must make a deposit into my bank account). And if this is true, we can see that Hobbes has not thoroughly considered the implications of his remarks about coercion for actual legal regimes. For it might be possible to forbid the use of coercion but nevertheless enforce

contracts flowing from that coercion.

The foregoing conclusion is bad news for Hobbes. For it implies that rational agents are *not* in fact forced to choose between a regime in which coercion is permitted in every context and one in which it is prohibited across the board. There may be many more possibilities from which Hobbesian agents can choose. And if they have a greater array of options, and if the rational incentives are as I suggested, then Hobbesian agents might well prefer a regime that enforced at least some coercive contracts. The simple argument I presented above about reversion to a state of nature would then be unavailable to a Hobbesian.

In particular, there is an intermediate regime that might prove attractive to Hobbesian agents: illegalize the initial use of coercion, attempt to obtain all required deterrent efficacy from that interdiction, yet nevertheless uphold agreements based on the use of illegal coercion. Such an approach is at least a logical possibility. An analogy might be the IRS's position on income from illegal activities: it is illegal to make money by smuggling drugs, but the duty to pay taxes on money so earned still exists, and it is a separate offense to fail to pay taxes on such earnings. Another example might be the relationship between trespassing and adverse possession: it may be illegal to trespass on someone's land, but if one does so consistently (open and notoriously) for twenty years, one acquires an easement to the land where one formally trespassed. It would clearly be *possible*, though not necessarily desirable, for a state to illegalize the use of coercion and still enforce contracts that emerge from illegal methods of bargaining.⁸⁸

Does it seem likely that the individuals attempting to maximize their own utility in the context of bargaining over an initial social contract would end up favoring a mixed regime of this sort—the regime that outlaws coercion but enforces coercive contracts once coercion is nevertheless illegally used? Such a conclusion seems possible on Hobbesian premises.

Outlawing the use of coercion is a perfectly sensible thing for a sovereign to do by Hobbes' lights, as it would discourage the use of coercion and thereby make individuals more secure. And outlawing only the initial use of coercion preserves the victim's ability to buy her way out of coercive encounters, which preferences we are bound to respect. In particular, if the primary coercion can be separately outlawed, there seems no justification for invalidating "voluntary" agreements that represent its downstream effects if the only reason to do so is to further discourage instances of primary coercion.

We can now connect this point with my original suggestion about the preferences of the victim. Notice that invalidating coercive contracts forces the victim of coercion to pay for some of the costs of future deterrence, *as opposed to keeping all deterrent costs firmly on the perpetrator*. This is because the victim is now deprived of the chance to save herself by striking a deal with the robber, in order to deter future coercers from committing similar acts. But arguably it would be more equitable, as between coercer who is in the wrong and victim who has done nothing wrong, to put all the burden of deterrence on the coercer and sharply increase his penalty for the initial use of coercion, thus exempting the victim from having to pay with her life for the benefit of overall social utility. And as arguably if such an arrangement would be more equitable, it would be more likely to be chosen by the parties as their preferred social arrangement, given that they do not know their positions *ex ante*.

Someone might object to this conclusion on the grounds that the coercer also gains if the agreement with his victim is enforceable. But that benefit can be overcome by simply further increasing the punishment for the initial coercion, since there is no particular difficulty with piling on ever increasing penalties for the initial coercive act.⁸⁹ So to be sure, if coercive contracts are enforced, the penalties for coercion will need to be increased. But this hardly seems

an objection, given all that has been said. It was then this mixed response that Hobbes did not, but should have, considered given other aspects of his philosophy.

Hobbes may now, however, be boxed into a corner. I argued previously that Hobbes is not as out of synch with common moral intuitions as one might have supposed upon first hearing his suggestion that the agreement with the robber is binding in a state of nature, for it is clear that he is prepared to invalidate coercive contracts if there is reason to legalize coercion generally in civil society. I then noted, however, that Hobbes appears to have missed the fact that coercion can be separately punished, and so it is in theory possible to enforce a contract that results from an illegal use of coercion. And in noting this possibility, Hobbes is deprived of an argument for the common sense solution, which invalidates a contract if the means of obtaining it are themselves illegal. Despite the fact that enforcing the coercive contract in many situations has its attractions from the standpoint of the victim, then, it will seem to many a *reductio* of the Hobbesian approach to legal questions if it ends up endorsing contracts with robbers entered into under duress. Does Hobbes have a different way out, other than that we considered above?

Let me step back from the gunman scenario for the moment and return to this example a bit later. I will be in a better position to address it if I first consider a situation in which enforcing a coercive, or at least quasi-coercive contract is less objectionable than it appears to be in the gunman case. I turn instead to the related, but milder cases of contract renegotiation.

V. MORE PLAUSIBLE APPLICATIONS: EXPLOITATION AND THE PRE-EXISTING DUTY RULE

The so-called “pre-existing duty rule” in contract doctrine maintains that a promise or performance is not valid consideration for a contract if the party offering it already had a pre-existing duty to do the thing promised or performed.⁹⁰ Thus I cannot offer my boss my

willingness to use best efforts in exchange for increased salary if using such efforts was already among the conditions of my employment. Nor can I demand more pay in exchange for a job I have already contracted to do, if doing that job was already an obligation I had taken on for lesser pay under a previous contract. In such cases, standard doctrine has it, the new promise or performance is invalid because it is barred by the existence of previous consideration.⁹¹

The pre-existing duty doctrine has traditionally been explained in terms of the doctrine of consideration: a commitment to do something one already has a duty to do cannot supply the consideration for a new obligation on the other party's part because there is nothing new to be exchanged. The doctrine of consideration is an expression of the requirement of exchange. Parties cannot bind themselves to a promise unless their promise is offered in exchange for the promise (or performance) of another, the benefit of which the party offering the promise hoped to obtain.

The explanation most frequently offered for the pre-existing duty rule, however, does not have to do with consideration but rather with exploitation.⁹² The concern is that without the rule, one party would be in a position to exploit the other midway through performance and demand increased wages or other benefits. In the absence of the rule, the other party would be forced to agree to the increase, as he would find himself in a particularly difficult situation if deprived of all means to induce the other to complete his work. It would seem, then, that having a pre-existing duty rule helps to avoid a "hold-up" problem that would occur in many contractual relations. The real rationale for the rule therefore, has to do with coercion, rather than consideration.⁹³

Contracts scholars, however, have noted that this rule can work to the disadvantage of both parties.⁹⁴ Thus in the famous *Alaska Packers* case, fishermen threatened to walk away from

their employment at a cannery in the middle of a fishing season, unless they received higher wages than originally agreed.⁹⁵ In response to the threat of losing its workforce midseason the cannery promised to pay the higher wages, and the fishermen remained on the job. But when the cannery failed to pay on the renegotiated contract at the end of the season, the fishermen sued, and the court ruled that the cannery's subsequent promise was barred by the prior duty the fishermen had to work for their original wages. The renegotiated contract was found to be barred by the pre-existing duty rule.

Suppose, now, that both sides had been more aware of the pre-existing duty rule than it appears they were. Would either side have been pleased to know at the time of renegotiation that any new agreement they reached would likely be barred in this way? As in the robbery situation, it is clear that both parties would have preferred *not* to have a pre-existing duty rule at the time of renegotiation: the fishermen clearly would have preferred that, since they were hoping to obtain a commitment from the cannery to pay higher wages. But the cannery too—like the victim in the robber case—would also have preferred to be able to renegotiate the contract, given that in the absence of credible renegotiation, the fishermen would have walked off the job. So just as in the actual hold-up situation, the fishermen and the cannery might both prefer *not* to be barred from renegotiating a binding contract at the moment of renegotiation.

Notice that this answer depends on several conditions being satisfied. First, as Oren Bar-Gill and Omri Ben-Shahar have emphasized, the threat to breach the contract must be a credible one.⁹⁶ If the fishermen had threatened to breach the contract and the cannery had reason to doubt their sincerity, the cannery might have preferred to have the support of the pre-existing duty rule in refusing to accede to their demands. Second, there must be the legal equivalent of the *common knowledge of rationality*, which we might refer to as the *common knowledge of legality*. Under

this condition, both parties have knowledge of the prevailing legal rules, and each has knowledge of each other's knowledge of such rules. Under these conditions, in the face of a sincere threat to breach the contract, the cannery would prefer not to have a pre-existing duty rule than to have it, as long as it can be assumed to be clear to both that the renegotiated contract will be binding. It is, after all, that advance knowledge on which the cannery relies, and any uncertainty on the fishermen's part will undermine the cannery's ability to convince the fishermen of the sincerity of its promise to pay.

Of course, the fact that the parties at the *moment of renegotiation* would want to abandon the pre-existing duty rule is neither surprising nor conclusive, given what we have seen in the case of the robber. The relevant question instead is whether they would have preferred to leave open the possibility of contract renegotiation at the time of contracting to allow for revision, even in the absence of changed circumstances. A rational conclusion would be that neither side would wish to take the risk of being subject to the other party's exploitation, and would want to guard against it. But consider the risks on the other side: there is the considerable danger of being on the receiving end of a credible breach of contract and having no means of convincing the other party to make good on its promise.⁹⁷ So the risks associated with not being able to engage in contract renegotiation may in the final analysis seem greater than the risk of exploitation, and both sides are likely to choose not to include the pre-existing duty rule in their initial contract.

In the case of contract renegotiation in the face of a sincere threat of breach, there is a maximally strong argument for enforcing the contract, despite the fact that the contract emerges from coercive, or exploitative tactics. Because of cases like this, legal scholars are increasingly calling for the elimination of the pre-existing duty rule.⁹⁸ Such scholars recognize the risk that those with greater bargaining power may try to exploit their position in the absence of the rule.

Yet most scholars who propose eliminating the pre-existing duty rule predicate their suggestion on the existence of another important background legal rule: the defense of duress. We could therefore eliminate the pre-existing duty rule as long as we retain the duress defense, or so they say.⁹⁹ The suggestion of such scholars thus seems to be that we should give effect to coercively obtained contracts in a renegotiation situation, as long as the coercion is not too extensive. Once the coercion rises to the level of true duress, we should continue to punish the coercer by invalidating the resulting contract.

While the intuition is a sound one, there is a problem with this approach, as I have already implicitly shown: the duress defense is subject to the same argument that compels the above scholars to abandon the pre-existing duty rule: both parties might see themselves as better off and neither worse off under particular agreements the defense forbids. Duress is in this regard indistinguishable from the pre-existing duty rule, and the two appear to be separated only by the *degree* of coercion at issue. Just as the coercion exercised by the fishermen against the cannery does not eliminate the cannery's desire to enter into a binding renegotiated contract for higher wages, so the robber's coercive tactics against his victim do not eliminate the victim's desire to bind himself to an enforceable agreement to pay the robber to release him. The coercion on which the duress defense is predicated, in other words, does not eliminate the preference for binding commitments. On the contrary—it appears to increase it. So while it seems in many circumstances as though the renegotiation should be allowed, the argument for eliminating the pre-existing duty rule is subject to the *reductio* that it would result in enforcing deals between robbers and their coerced victims. What we need to defend the common sense solution, then, is a way of distinguishing the gunman situation from the renegotiation context. That is, we need a way of preserving the duress defense in the face of a sensible rejection of the

pre-existing duty rule.

Let us return to the suggestion of the mixed regime considered above. Recall that this suggestion seemed attractive from the standpoint of rational choice considerations, since it would achieve the deterrent benefits hoped for by invalidating coerced contracts, at the same time that it would relieve the victim from bearing some of the costs of that deterrent efficacy. It is also clear, however, that the suggestion is likely unacceptable as an actual, administrative response to the gunman scenario, and so its availability merely served in my argument to detract from, rather than enhance, the attractiveness of the Hobbesian approach. The same solution, however, is vastly more appealing in the context of the pre-existing duty rule. That is, we could quite sensibly enforce the renegotiated contract between the cannery and the fishermen, yet sanction the fishermen independently for any improper or coercive tactics they might have used in gaining the agreement of the cannery. How would this kind of solution be implemented?

A court would first have to determine whether the fishermen had indeed behaved coercively in threatening to walk off the job. If they were responding to a real economic necessity—for example, they were underpaid and discovered they could benefit significantly by taking employment elsewhere—it is less likely that the threat to breach the contract would be seen as coercive. If the threat to breach the contract were a bluff or otherwise poorly motivated, a court might find coercion, and set a penalty. The size of the penalty would be a reflection of the degree of coercion the court believed was involved. Thus although the renegotiated contract would be allowed to stand, since it was agreed to voluntarily, the fishermen's award might be reduced by the amount of coercion they used to get the cannery to offer that agreement. In the event of extreme coercion, there would be no return to the fishermen, and might indeed be a further penalty (whether civil or criminal) extracted as punishment for improper behavior.

Now let us compare this bifurcated regime in the renegotiation context to the same approach in the gunman scenario. The problem is that since it is possible to punish the gunman separately for his coercion—we have a distinct crime, armed robbery, that is separately prohibited and so can be separately punished—why would we not be willing to say that the deal with the gunman should *always* stand, and that the general deterrence we require could in this instance be separately obtained by extensive punishment of the original coercion?¹⁰⁰ Indeed, at least one pair of legal scholars thinks no appropriate distinction can be drawn. Oren Bar-Gill and Omri Ben-Shahar have written that whether a coerced agreement should be enforced should turn on the credibility of the threat of the coercing party.¹⁰¹ A threat is credible if the coercer's payoff for carrying out the threat exceeds the payoff from not carrying the threat out.¹⁰² The idea, like that pointed out above, is that the possibility of an ex post remedy induces the coercer to carry out the threat since their payoff would be greater. When the coercer is bluffing, however, an ex post remedy could not induce the bluffing coercer to carry out the threat, given that they presumably saw threat-execution as not worthwhile in the first place.¹⁰³ Consequently, Bar-Gill and Ben-Shahar argue that whether an ex post remedy should be available for a coerced party should turn only on whether the threat was credible or a mere bluff.

That Bar-Gill and Ben-Shahar reach such counter-intuitive results is obviously problematic for their account. What is perhaps more problematic for them, however, is that the criterion of credible threat does not distinguish the serious case of the gunman from the less serious case of the *Alaska Packers*.¹⁰⁴ This is because in all cases the credibility criterion is a “descriptive understanding of the threatening party's incentives,” which determine whether the “ideal outcome” is attainable.¹⁰⁵ Credible threats, as well as bluffs, are equally possible in all cases of coercion, regardless of severity.

It looks as though the more *severe* the coercion the more likely we will be able to separate it from the resulting contract, and the more appropriate separate punitive measures would be, and hence the more justified we would be in enforcing the coerced contract! But this is the reverse of what intuition would tell us, since we are more willing to enforce the resulting contract in a mild case of coercion like *Alaska Packers* than in a deal negotiated at gunpoint. In order to reject this logic, let us consider whether there is not a way of distinguishing the logic of the gunman-type scenario from the pre-existing duty-type case, such that we can reach the more intuitive result, namely to employ the bifurcated approach in the latter case but not in the former.

There is one natural way to distinguish the cases, which I highlight only to reject. One might be tempted to distinguish the cases in terms of the severity of the threat. Thus, the fishermen in *Alaska Packers* are threatening to breach an antecedent contract between themselves and the cannery, where the relevant contract is one made between the “aggressor” and the “victim” at an earlier moment in time. By contrast, the robber threatens what would appear to be a much greater wrong than mere breach of contract: he threatens to kill the victim. Since killing is (usually) worse than the ordinary violation of a civil employment contract, we would ideally like to distinguish these two threats in terms of *degree of wrongfulness*, and say that the contract is disallowed in the gunman case because the threat from which it results is worse. But *degree of wrongfulness* would appear to return us to the deontological approach I rejected on rational choice grounds at the outset. We must therefore find a different way. Is there another basis for distinguishing the two types of threats?

Our problem, as is now clear, is that the two cases are structurally highly similar, indeed more similar than might first be thought. For the wrongfulness of unjustified killing in a Hobbesian system is also a function of contract violation, albeit one made not with the victim

alone, but with all of society. In a Hobbesian system, the way to explain the wrongfulness of killing is that unprovoked or non-defensive intentional killing violates the very terms of the social contract itself. It will therefore not work simply to say that killing is worse than contract violation and to distinguish the cases *that way*, since the wrongfulness of killing is itself inseparable from the notion of contract violation.

VI. THE SOCIAL CONTRACT AND COERCION

The observation that the robber's threatened action violates the terms of the social contract should now allow us to articulate a crucial distinction between the *Alaska Packers* case and the case of the deal with the robber. While it may be open to the victim to waive the terms of a contract that he has made with the coercer in the pre-existing duty context, the victim is *not* free to waive the contract violation where all other members of society are also party to the agreement. For in doing so, the victim would undermine not only his own entitlement to insist on fidelity to mutually beneficial exchanges, but the right of everyone else to do the same. Thus what the robber is threatening to do, should the victim fail to agree to the offered contract, is to violate the defensive conditions of a contract entered into with all other members of society, and this is something the victim is not entitled to do. Waiving defensive rights of the social contract would render the position of all insecure by eliminating the protections they have chosen for their agreement, and *that* it is not the prerogative of any given individual.

This of course explains why matters would be different in a state of nature. There the robber's threat is not "wrongful," since there is no social contract the robber is threatening to violate. And this is just to say that the robber can use the benefit of his threat advantage in a state of nature, but not in civil society. The point of the social contract after all, is to rule out

precisely that sort of threat advantage.

An objection might be raised that the robber is only *threatening* to breach the contract—he is not actually breaching the contract, unless the victim fails to comply with his request. While there are different positions among philosophers about the relationship between threats and actions in furtherance of such threats, it is clear from the literature that it is a perfectly reasonable position to suppose that threatening to violate a contract is the same wrongful act as actually *violating it*. Admittedly, threatening to do what is wrongful might be its own kind of wrong, distinguishable from the wrong of the threatened action. It might, for example, be a wrong of a lesser order, related of course to the wrong that is itself threatened, but still a wrong of a different, and probably lesser, significance. But I have not argued that threatening to violate the contract is itself a violation of the social contract, and it is indeed unclear whether such a wrong should be ranked as among the wrongs that the parties would be concerned to protect themselves against in the basic defensive social covenant. So the focus should probably not be on the potential wrongfulness of the threat itself, but rather on the wrongfulness of the action being threatened. And the question about *that* action must be whether, by entering into an agreement to pay the robber for his suspension of violence, the victim is effectively releasing the robber from his commitment to the social contract. For this is what I have claimed the victim may not do.

It is reasonably clear that he is. First, there is the relatively straightforward fact that the robber will surely not enter the agreement unless it contains an immunity from prosecution in addition to the promise to pay. For if the robber is to be prosecuted, he will almost certainly fail to reap the gains he was hoping for under the agreement. And to ask for immunity from prosecution is effectively to ask for an exemption from the consequences of violating the social

contract, the possibility of which we can understand the robber has already rejected. Even the victim does not have the power to grant immunity from prosecution. Criminal prosecution is a public function, a fact that underscores the idea that it is a wrong of a different sort that is being threatened here. The wrong of killing the victim is not only a wrong to the victim; it is a wrong to society as a whole, and therefore society retains the right to prosecute it. This right—the right of prosecution under the social contract—is one the victim is powerless to waive.

Thus the rational choice analysis of contracts under coercion suggests leaving the invalidation of contracts intact in cases in which the initial contract is entered into under coercion, but allowing the contract to stand where the coercion is employed in the context of contract renegotiation. This approach would honor the fact that individuals acting under voluntary coercion are not *literally* deprived of choice. They actually do exercise choice, and there are surely many cases in which their welfare may depend on their ability to enter into the very bargain they would later like to invalidate. But because this suggestion would have untoward effects on the social contract in cases of true coercively induced initial agreements, we invalidate the contract in all such cases, despite the immediate benefits to the victim from bargaining with gunmen.

VII. A FURTHER TEST: PLEA BARGAINING AS CONTRACT

Let us turn now to a different application of the bifurcated approach to coercive contracts, one that appears in the law enforcement setting. Much has been written in recent years about the voluntariness of agreements reached as part of a plea bargain. And indeed many legal scholars have objected to plea bargains on the ground that such agreements are the product of coercion.¹⁰⁶ They maintain that if we were applying ordinary contract principles to these cases, plea

bargaining, agreements would rightly be void. For example, a prosecutor who threatens a criminal defendant with prosecution for capital murder to gain the defendant's agreement to a twenty year sentence is thought to be like the gunman in the initial example: the defendant must effectively agree to plead guilty and accept the sentence under pain of worse alternatives. The choice he gives the defendant is said to be no choice at all; hence the resulting plea cannot be voluntary, it is argued.

Just as I have shown in the case of ordinary agreements and contracts, this description of the agreement between prosecutor and defendant requires some careful parsing. The plea bargain, unless physically forced or agreed to under torture, is surely fully volitional in the sense discussed earlier. And lawyers do not have an alternative definition of voluntariness to put in its place. According to Supreme Court jurisprudence on this question, for example, a plea is "involuntary" when "the defendant's will is overborne,"¹⁰⁷ when the defendant is not in "possession of his mental freedom,"¹⁰⁸ when "the accused was deprived of his free choice to admit, deny, or refuse to answer,"¹⁰⁹ when "the choice is between the rock and the whirlpool,"¹¹⁰ and so forth. But it seems substantially more likely that the relevant feature of such situations that courts are trying to capture has nothing to do with lack of voluntariness, as I argued earlier in the case of agreements of other sorts.

The reason that the jurisprudence of plea bargaining has developed this confused notion of voluntariness is clear enough. Courts, like lawyers and philosophers, make the mistaken assumption that unless we are prepared to say that a person gave his assent to an agreement *involuntarily*, we will have no basis for voiding the agreement. Thus the move is made to posit a psychological condition under which the defendant is not responsible for his choices, and then to label the agreement made in that condition "coercive," in order to attack its enforceability. But

the flaws in this line of reasoning should by now be clear, given what we have seen in the case of private contracts. First, there is no reason to suppose that we must treat an agreement as involuntary in order to call it coercive. Second, the jurisprudence of plea bargaining assumes that unless a plea is dubbed “coercive,” we would have no basis for rejecting it. But as in ordinary contracts, neither of these assumptions is justified. Indeed, in some cases, we may well wish to say that the defendant was coerced into accepting a plea bargain, without either regarding his assent as involuntarily offered or invalidating the plea.

In particular, suppose we generally regard pleas as acceptable, but we wish to invalidate them when they are produced by certain tactics or methods on the part of prosecutors. And we believe they should be invalidated in such cases, despite the fact that we regard them as fully voluntary. In *this* case, voiding the agreement is unlikely to be an effective deterrent of coercive tactics on the prosecutor’s part, because the prosecutor will find himself in precisely the situation he would have been in had he never initiated negotiations with the defendant: he will have the choice to bring the case to trial or to drop it, or perhaps to explore new pleas that are likely to be upheld. But he is not personally penalized for his improper behavior, at least in most cases. The bifurcated approach to coercion might be particularly appropriate under this set of circumstances: a defendant who was coerced into a plea agreement could be held to that agreement, if his guilt was reasonably supported by additional evidence, other than the plea. At the same time, however, the prosecutor could be separately sanctioned for the use of coercive tactics. While there are of course criminal sanctions available for prosecutors and police officers who go *far* beyond the limits of their offices, the threshold for sanctioning police and prosecutors for improper techniques for achieving pleas is set quite high as matters presently stand. The bifurcated approach would recommend setting the threshold substantially lower, but sanctioning

prosecutorial misconduct directly.¹¹¹

In conclusion, it is worth noting that there is a crucial, substantive reason to be concerned about the practice of plea bargaining that has nothing to do with voluntariness or coercion, and this is the threat that plea bargaining poses to the presumption of innocence. What is worrisome is that defendants are induced by the practice of plea bargaining to abandon their constitutional right to prove their innocence by the fear of what will happen if they are unable to succeed at trial. Detractors of the practice once again have a tendency to say that defendants are “coerced” into giving up their right to a jury trial. But offers of plea agreements need not be thought any more coercive than many attractive offers might be. Furthermore, we can object to the inducement the plea provides without calling the resulting agreement “coercive,” or, alternatively, without seeking to invalidate it for this reason. The objection may simply be that we do not want defendants giving up their right to trial by jury *for the wrong reasons*. In particular, we do not want *innocent* defendants to be induced to give up that right, since their reasons for giving up the right surely do not advance the values the criminal justice system seeks to endorse.

I will not explore this line of argument further on the present occasion, as it is not directly relevant to my immediate concerns. What should be clear is that in the plea bargaining context, as in other areas of the law, an important, substantive ethical and political concern has been obfuscated by the language of involuntariness and its mistaken association with coercion, as well as by confusion about the implications of labeling an agreement or an action as coercive. The motivation for the obfuscation is the difficulty itself of reaching desired ethical conclusions with straightforward moral reasoning. Once we recognize that the mistaken psychologizing of coercion is partially motivated by ethical concerns, it should help us to disentangle the ethical

from the psychological and start afresh.

VIII. CONCLUSION

In this Article, I have sought to expose a tension between the perspective of a victim in cases of coerced contracts and the usual justifications that are offered for invalidating such contracts. In particular, I have suggested that the victim in such cases would regard refusing to enforce such contracts as a kind of double insult to her well-being: first the coercer lowers her baseline welfare by forcing her to choose between her life and her economic well-being, and then the state lowers it even further, by refusing to allow the victim to make the binding commitment that would enable her to make the best of her remaining options. I have also argued that an adequate justification for a social policy of this sort would require that the policy be justified *to the victim*, in terms that the victim can endorse and regard as fair to her interests. Such justifications would be contractarian in nature, and would differ substantially from the types of justifications typically on offer, which tend to be moralized or consequentialist.

In the attempt to discover a justification for barring coercive contracts, I noted that such contracts strongly resemble a situation involving the use of the pre-existing duty rule, in which parties attempt to renegotiate contracts under threat of breach from the other party. Often they discover that they cannot engage in such renegotiation because it would be barred by the other side's pre-existing duty to meet his contractual commitments. In such cases, intuition strongly supports allowing the parties to engage in binding renegotiation, despite the pre-existing duty rule. And this intuition is strongly supported by the contractarian thought that it is unfair to bar the victim to buy her way out of a coercive situation by making binding agreements under the pressure of some adverse threat. If the contractarian argument supports enforcing agreements

reached under threat of breach of contract, despite the “coercive” aspect of such agreements, why would it not support enforcement of contracts made with gunmen? As I noted, at least some contracts scholars in the economic tradition have reached the conclusion that the two situations are identical in this respect. And it is true that from an economic perspective, there is little to distinguish the cases. But intuition speaks strongly in favor of trying to draw such a distinction, and as it happens, such a distinction is supported by contractarian reflection.

I have argued that it makes sense to think that the rights that are being waived in the pre-existing duty case are rights the victim entirely controls: the norm being violated in that case is the contractual commitment that the perpetrator previously made to the victim. In the gunman hypothetical, however, the norms being violated are general background norms having to do with the collective commitment to the inviolability of human life. These the victim is powerless to waive. I therefore conclude that a contractarian approach can plausibly distinguish cases of coercion from one another, and that in some instances, the fact that the contract is entered into under duress, or quasi-duress, does not entail its invalidation. In others, however, the existence of duress *does* have that implication, and there the traditional wisdom that coercive contracts are invalid is precisely right. As we have seen, however, this is not for the reasons that traditional analysis would once have provided.

Notes

¹ The *Restatement (Second) of Contracts* § 175 defines duress as follows:

If a party's manifestation of assent is induced by an *improper threat* by the other party that leaves the victim *no reasonable alternative*, the contract is voidable by the victim.

American Law Institute, *Restatement (Second) of Contracts*, § 175 (emphasis added). Compare this with the *Model Penal Code*'s definition, establishing duress as a defense to a crime:

It is an affirmative defense that the actor engaged in the conduct charged to constitute an offense because he was coerced to do so by *the use of, or a threat to use, unlawful force* against his person or the person of another, that a *person of reasonable firmness in his situation* would have been unable to resist.

American Law Institute, *Model Penal Code*, § 2.09 (emphasis added). The criminal law is concerned with coercion through *unlawful force* rather than an *improper threat*. The class of improper threats is clearly larger than the class of unlawful threats. Moreover, the *Model Penal Code*'s focus on the objective standard of the "person of reasonable firmness" is distinct from contract's criteria of "no reasonable alternatives." It is unclear, however, whether there is divergence between the two criteria. Presumably, one way to assess the reasonableness of one's alternatives is to think about whether a person of reasonable firmness would have no alternative.

² Default rules are rules that courts will use to interpret contracts unless the parties have specified to the contrary. Mandatory rules, by contrast, are rules of contract formation that the parties are not free to ignore. See, e.g., E. A. Farnsworth, *Contracts*, 4th ed. (Aspen, 2004), § 7.15-.17.

³ See *Dickinson v. Dodds*, L.R. 2 Ch. D. 463, 472 (1876) ("It must, to constitute a

contract, appear that the two minds were at one”) (James, L.J).

⁴ “The rules on duress . . . allow the injured party to undo the transaction by avoiding it.”
Farnsworth, *Contracts*, § 4.19.

⁵ Charles Fried, *Contract as Promise: A Theory of Contractual Obligation* (Cambridge, MA: Harvard University Press, 1981), 98-99 (“I conclude that a promise procured by a threat to do wrong to the promisor . . . is without moral force.”); Alan Wertheimer, *Coercion* (Princeton: Princeton University Press, 1987), 3 (“A coerced promise is not morally binding.”).

⁶ See, e.g., Joel Feinberg, *Harm to Self* (New York: Oxford University Press, 1984), 189-268; Harry Frankfurt, “Coercion and Moral Responsibility,” in *Essays on Freedom of Action*, ed. T. Honderich (London: Routledge and Kegan Paul, 1973), 63-86; Charles Fried, *Contract as Promise: A Theory of Contractual Obligation* (Cambridge, MA: Harvard University Press, 1981), 93-98; Daniel Lyons, “Welcome Threats and Coercive Offers,” *Philosophy* 50 (1975): 425; Robert Nozick, “Coercion,” in *Socratic Puzzles* (Cambridge, MA: Harvard University Press, 1997); Wertheimer, *Coercion*; David Zimmerman, “Coercive Wage Offers,” *Philosophy & Public Affairs* 10 (1981): 121.

⁷ See, e.g., Oren Bar-Gill and Omri Ben-Shahar, “Credible Coercion,” *Texas Law Review* 83 (2005): 717; Richard A. Posner, *Economic Analysis of Law*, 4th ed. (Boston: Little, Brown, 1972), 115. For parallel implicative discussions of duress in the substance of the criminal law, see Claire Finkelstein, “Duress: A Philosophical Account of the Defense in Law,” *Arizona Law Review* 37 (1995): 281, and in the plea-bargaining context see Oren Bar-Gill and Omri Ben-Shahar, “Plea Bargains Only for the Guilty,” *Journal of Law and Economics* 49 (2006): 353; Kenneth Kipnis, “Criminal Justice and the Negotiated Plea,” *Ethics* 86 (1976): 93; Stephen J. Schulhofer, “Plea Bargaining as Disaster,” *Yale Law Journal* 101 (1992): 1979; Robert E. Scott

and William J. Stuntz, "Plea Bargaining as Contract," *Yale Law Journal* 108 (1992): 1918-24. Finally, there are "implicative" explorations of the related topic of the pre-existing duty rule. See Oren Bar-Gill and Omri Ben-Shahar, "Credible Coercion"; Randy E. Barnett, "A Consent Theory of Contract," *Columbia Law Review* 86 (1986): 269; Robert L. Hale, "Bargaining, Duress, and Economic Liberty," *Columbia Law Review* 43 (1943): 603.

⁸ These categories are rough and ready. For example, one legal scholar discusses the concept, or rather "concepts," of coercion in a way that does not seem to fit neatly into the categories I have described. Mitchell Berman thinks that the appropriate way to explore coercion is to look at how "coercion claims" function in our practical reasoning and to think about the types of questions that coercion claims answer. On this approach he thinks that there are different concepts of coercion that generate different implications, for example, whether a coercion claim counts as an excuse, in different contexts. So, the account is definitional in that it identifies different kinds of coercion but implicative in that these the definitions are connected with specific implications. Mitchell N. Berman, "The Normative Functions of Coercion Claims," *Legal Theory* 8 (2002): 4.

⁹ Since my interest is not in offering an account of coercion but in the question of the appropriate treatment of coercive agreements, I shall not be excessively concerned with the "definitional" questions on which most philosophers focus, but will explore them only to the extent required to address the "implicative" question.

¹⁰ It turns out, however, that not all possible combinations are extensively represented in the literature, and that some writers on coercion might be thought to take a mixed approach, which may, for example, be focused on the definitional task but solely for the purpose of its implications. This is the approach that I understand Mitchell Berman to be taking. Berman,

“Normative Functions.” Other thinkers rely on an implicit definition of coercion but focus on a criterion by which to identify not which cases are cases of coercion, but rather which cases should be dealt with under the law. Oren Bar-Gill and Omri Ben-Shahar use their “credibility criterion” to serve this function. “Credible Coercion.” Finally, some views are structured to leave open the particular moral theory to which the account of coercion needs to be tied. In Wertheimer’s account of coercion baselines, the moral theory that sets the moral baseline theoretically can be left open, and perhaps be utilitarian, despite Wertheimer himself being a deontologist. Wertheimer, *Coercion*, 212.

¹¹ This is somewhat misleading because the psychological accounts are also somewhat moralized in the sense that they are normative, yet their normative claims rest on psychological or empirical premises rather than moral ones, and if this is borne in mind, these categories should not confuse.

¹² *Miller v. Fenton*, 474 U.S. 104, 116 (1985).

¹³ *Lyons v. Okla.*, 322 U.S. 596, 601 (1944).

¹⁴ See Wertheimer, *Coercion*, 8 (“[C]oercion negates or compromises the voluntariness of an act . . .”). Wertheimer, however, is better classified as a moralized deontologist, as will be clear below.

¹⁵ For a criticism of the psychological approach, see Claire Finkelstein, “Duress: A Philosophical Account of the Defense in Law,” *Arizona Law Review* 37 (1995): 281.

¹⁶ Paul H. Robinson, *Criminal Law Defenses* (St. Paul: West, 1984), 1:351.

¹⁷ George P. Fletcher, *Rethinking Criminal Law* (New York: Oxford University Press, 2000), 811 (emphasis added).

¹⁸ See Wertheimer, *Coercion*, 30-31. To be more precise, Wertheimer characterizes the choice prong in terms of what would count as not a “reasonable choice,” rather than whether there was any opportunity for choice at all. *Ibid.*, 36.

¹⁹ *Ibid.*, 201-02. Wertheimer believes that nothing can come of an account that determines choice based on the desirability of the set of alternatives one faces or the relative desirability (what Wertheimer calls “distance”) of the alternative(s) when compared with other alternatives in the set. *Ibid.*, 192-201. Rather, Wertheimer regards the essential element in coercion as an impairment of the degree to which the victim’s choice was a manifestation of his second-order preferences. *Ibid.*, 202 and 301-06. Wertheimer believes his two-prong analysis tracks the Second Restatement’s two duress provisions. *Ibid.*, 30. See *supra* note 1.

²⁰ *Ibid.*, 38-46.

²¹ *Ibid.*, 4.

²² *Ibid.*, 217 (“[T]he structure of coercion discourse presupposes that A and B have certain obligations and rights which establish a background against which A’s proposals are understood.”)

²³ *Ibid.*, 217-221. While it is clear that Wertheimer’s account centers around his account of moral baselines, he is often non-committal with respect to whether a moral test is the always the correct test for determining a potential victim’s baseline. He states: “There is no single right answer . . . when the results of the nonmoral and moral tests [for baselines] diverge. Each test supports a defensible or plausible coercion claim.” *Coercion*, 212.

²⁴ I call this view “counterintuitive,” but it incorporates a fairly standard move among philosophers, namely to think of acting against second order principles one holds as acting involuntarily. See, e.g., Harry G. Frankfurt, “Freedom of the Will and the Concept of the

Person,” *Journal of Philosophy* 68 (1971): 5-20.

²⁵ See Nozick, “Coercion,” 6-22.

²⁶ *Ibid.*, 24.

²⁷ Wertheimer, explicitly leaves it open: “Whether these moral requirements are *ultimately* grounded in a deontological or consequentialist theory, the structure of coercion discourse presupposes that A and B have certain obligations and rights which establish a background against which A’s proposals are understood.” *Coercion*, 217.

²⁸ Richard A. Posner, *Economic Analysis of Law*, 4th ed. (Boston: Little, Brown, 1972), 115. See also Steven Shavell, who notes that “voiding removes the socially wasteful incentive to engineer situations of need.” “Contractual Holdup and Legal Intervention,” *The Journal of Legal Studies* 36 (2007): 336. Later I discuss Oren Bar-Gill and Omri Ben-Shahar in contrast with my own view. It is worth mentioning that while their view on coerced contracts is similar in many respects to a perspective I take here, it diverges in its dependence on consequentialist concerns. For example, Bar-Gill and Ben-Shahar worry that enforcing coerced contracts might encourage “investments” in soliciting harmful information for purposes of blackmail and encourage highway gunmen to make themselves more credible coercers, a consideration that would not be directly relevant in my view. “Credible Coercion,” *Texas Law Review* 83 (2005): 732-33. However, their view seems to diverge from a completely consequentialist approach in that they, like myself, are primarily concerned with the well-being of the potential victims of coercion. *Ibid.*, 718.

²⁹ Robert Cooter and Thomas Ulen, *Law and Economics* (Harper Collins, 1988), 254.

³⁰ *Ibid.*, 252.

³¹ *Ibid.*

³² Ibid., 254.

³³ See Claire Finkelstein, "A Contractarian Argument against the Death Penalty," *New York University Law Review* 81 (2006): 1308, for an overview of the distinction.

³⁴ See John Rawls, *A Theory of Justice* (Cambridge, MA: Belknap Press, 1971).

³⁵ See Thomas Scanlon, *What We Owe to Each Other* (Cambridge, MA.: Harvard University Press, 1998).

³⁶ Wertheimer would probably concede the point, in so far as he thinks the idea of voluntariness is a normative, not an empirical concept. *Coercion*, 290. And from a normative standpoint, talk about mental processes being short-circuited would be out of place.

³⁷ So when Hobbes says, "of the voluntary acts of every man the object is some *good to himself*," *Leviathan*, chap. XIV, par. 8, he need not be understood as endorsing psychological egoism. Instead he can be understood as falling into line with the traditional teleological account of intentional action, according to which an action is correctly described as "voluntary" just in case it is intentional, under some description, and it is intentional under a description just in case it is performed under that description for the sake of an end.

³⁸ Wertheimer, *Coercion*, 9. Non-volitional coercion would be a case in which one's actions are literally physically compelled.

³⁹ See Thomas C. Schelling, *The Strategy of Conflict* (Cambridge, MA: Harvard University Press, 1960).

⁴⁰ For a discussion of this point in the criminal law context, see Finkelstein, "Duress," 281.

⁴¹ Sometimes the line between volitional and non-volitional coercion may be difficult to draw, and this is an admitted difficulty with these categories. For example, is having one's hand

physically forced to sign a document so very different from being “forced” to sign a document at gunpoint? What about at knifepoint, where the point of the knife literally presses your hand in the desired direction? Distinctions with fuzzy edges are still distinctions, however, and the basic separation between volitional and non-volitional coercion seems nevertheless maintainable.

⁴² Nozick, “Coercion,” 24.

⁴³ Ibid.

⁴⁴ Ibid., 28.

⁴⁵ *Coercion*, 212-17.

⁴⁶ Posner, *Economic Analysis*, 115.

⁴⁷ But see John C. Harsanyi, “Morality and the Theory of Rational Behaviour” in *Utilitarianism and Beyond*, eds. Amartya Sen and Bernard Williams (New York: Cambridge University Press, 1982). Arguably a rational choice theorist like Harsanyi takes the individualistic perspective more seriously than other Consequentialists, despite the fact that his main purpose is to defend the principle of utility, on the grounds that individuals in an “original position” would choose that principle as their principle of distribution. It is beyond the scope of the present paper to discuss his view, since he himself does not discuss coercion. But it might be argued that a fuller economic account of the coercion literature could, nevertheless, make room for the individual perspective by embedding the principle of average utility in a bargaining framework. It is my own view, however, and the view of many other Contractarians, that the bargaining framework does not in fact lend support to the principle of utility.

⁴⁸ For a greater appreciation of the differences between economic and rational choice approaches to law, see Claire Finkelstein, “Legal Theory and the Rational Actor,” in *The Oxford*

Handbook of Rationality, eds. A. R. Mele and P. Rawling (New York: Oxford University Press, 2004).

⁴⁹ For further discussion of the point, see *ibid.* But *cf.* John Harsanyi *supra* note 48.

⁵⁰ But see note 28.

⁵¹ Thomas Hobbes, “De Cive,” in *Man and Citizen*, ed. Bernard Gert (Indianapolis: Hackett, 1991), 129.

⁵² Thomas Hobbes, *Leviathan*, chap. XIV, par. 27.

⁵³ *Ibid.*, chap. XIV, par. 20.

⁵⁴ To avoid cumbersome formulations I have been referring to *an exchange of promises*, but, as will be clear shortly, that exchange can also be “unilateral,” namely an exchange of a promise for a present performance.

⁵⁵ Hobbes thinks that promises made without exchange, or gifts, are not binding if made for the future (“I promise I will give you my cow tomorrow”). But he thinks that gifts where the promisor transfers a right to another party *now* are binding, even if some part of the performance remains for later. “I hereby transfer my cow to you *now*, to be delivered tomorrow,” is a binding commitment to deliver the cow tomorrow. The first category of promise is the one in which the element of exchange makes a difference. See Hobbes, *Leviathan*, chap. XIV, pars. 9-16.

⁵⁶ *Ibid.*, chap. XIV, pars. 27.

⁵⁷ *Ibid.*, chap. XX, par. 1.

⁵⁸ *Ibid.*, chap. XX, par. 1.

⁵⁹ *Ibid.*, chap. XX, par. 2.

⁶⁰ *Ibid.*

⁶¹ *Ibid.*

⁶² Ibid., chap. XXI, par. 3.

⁶³ Ibid., chap. XIII, par. 13.

⁶⁴ Ibid., chap. XIV, par. 18.

⁶⁵ This distinction is the distinction between unilateral and bilateral contracts.

Farnsworth, *Contracts*, § 3.4.

⁶⁶ Ibid., chap. XIV, par. 9 (“The mutual transferring of right is that which men call CONTRACT.”).

⁶⁷ Ibid. (“[O]ne of the contractors may deliver the thing contracted for on his part, and leave the other to perform his part at some determinate time after (and in the meantime be trusted); and then the contract on his part is called . . . COVENANT . . .”).

⁶⁸ Howard Warrender, *The Political Philosophy of Hobbes: His Theory of Obligation* (Oxford: Oxford University Press, 2000), 30.

⁶⁹ Gregory S. Kavka, *Hobbesian Moral and Political Theory* (Princeton: Princeton University Press, 1986), 304.

⁷⁰ David P. Gauthier, *The Logic of Leviathan: The Moral and Political Theory of Thomas Hobbes* (Oxford: Clarendon Press, 1969), 41.

⁷¹ Hobbes, *Leviathan*, chap. XIV, par. 11 (emphasis added). The fact that Hobbes says “may” here should not be taken to mean that a covenant *may* be one in which one party must perform, but rather that the person may or may not end up performing on the covenant.

⁷² Ibid., chap. XIV, par. 9. He distinguishes transferring a right to a thing from transferring the thing itself. The suggestion is that a direct exchange of goods or of money for goods would not constitute a contract, since there are no ongoing obligations after the transaction has terminated. This is consistent with the position that modern law takes on the topic of barter.

⁷³ Farnsworth, *Contracts*, § 3.4.

⁷⁴ *Ibid.*

⁷⁵ Hobbes, *Leviathan*, chap. XIV, par. 18.

⁷⁶ Farnsworth, *Contracts*, § 3.4.

⁷⁷ *Ibid.*, chap. XV, par. 1.

⁷⁸ *Ibid.*, chap. XV, par. 4 (describing the fool as an individual who does not believe in justice).

⁷⁹ *Ibid.*, chap. XV, par. 5.

⁸⁰ *Ibid.*

⁸¹ I argue that Hobbes was committed to the common knowledge assumption as a feature of the state of nature in Chapter Two of my forthcoming manuscript. See “Contracts and Covenants” in *Hobbesian Legal Theory* (unpublished manuscript).

⁸² Cristina Bicchieri, *Rationality and Coordination* (New York: Cambridge University Press, 2007).

⁸³ This should not be misunderstood as suggesting that the robber’s willingness *causes* it to be rational for you to return with the money. It is rather that the superior position both you and the robber enjoy from the exchange makes it rational for each of you to conform your actions to the plan, *provided that the other is willing to do so as well*. Thus the robber’s willingness to abide by the plan signifies that he is acting in conformity with the rational course of action for both of you, and so suggests it is rational for you to do so as well.

⁸⁴ Wertheimer, *Coercion*, 238.

⁸⁵ Hobbes, *Leviathan*, chap. XX, par. 2.

⁸⁶ Example drawn from the same text in the Latin version.

⁸⁷ In contract law this might be known as the distinction between substantive and procedural unconscionability. Duress, of course, refers to the conditions under which the agreement is created. Thus it is largely a procedural rather than a substantive notion.

⁸⁸ There are surely cases in which it is difficult even to make conceptual sense of this distinction. Consider attempting to distinguish the initial coercion used in rape from the resulting illegal act of coerced intercourse. One can hardly make sense of the suggestion that the force might be illegal but the resulting sexual encounter legally protected.

⁸⁹ There is no upward limit to the amount of punishment that can be imposed for any act, as ever increasing amounts of torture can always be added to death. But there *is* potentially a problem with punishment bunching as punishments increase. This is the familiar problem that if the punishment for bike theft is increased in an effort to more strenuously deter this sort of crime, then the punishment for bike theft and for auto theft may turn out to be too close. In this case, the schedule of punishments will end up encouraging all potential bike thieves to steal automobiles instead, and that would not be an advantage. So if the penalty for the robber's initial act of coercion is sufficiently high, one risks encouraging greater, more lucrative crimes, etc. On the other hand, if there is no upward limit on punishment generally, then other punishments can in theory be adjusted upwards as needed.

⁹⁰ See, e.g., *Stilk v. Myrick*, 170 Eng. Rep. 1168, 1168 (1809) (L.R.C.P) (Ellenborough, L) (holding a renegotiated contract void due to lack of consideration).

⁹¹ Though the *Restatement (Second) of Contracts* says that the modification will be allowed if the change is "fair and equitable." § 89(a).

⁹² See Steven Shavell, "Contractual Holdup and Legal Intervention," *Journal of Legal Studies* 36 (2007): 334-35.

⁹³ Of course the doctrine of consideration itself may be seen as a way of guarding against coercion, since the existence of consideration suggests reciprocation, and hence fair exchange. But consideration in fact need not be fair in order to count as valid, and the doctrine is better understood as a way of distinguishing exchanges from bargains than as a substantive test for the fairness of the agreement.

⁹⁴ See Oren Bar-Gill and Omri Ben-Shahar, "The Law of Duress and the Economics of Credible Threats," *Journal of Legal Studies* 33 (2004): 391.

⁹⁵ See *Alaska Packers v. Domenico*, 117 F. 99 (9th Cir. 1902).

⁹⁶ "Threats known to be noncredible cannot and will not coerce." Oren Bar-Gill and Omri Ben-Shahar, "Credible Coercion," *Texas Law Review* 83 (2005): 727 (italics omitted).

⁹⁷ This is particularly so if the contract calls for personal service of some sort. Most courts will not order specific performance on such a promise because of worries about degrading servitude. See, e.g., *The Case of Mary Clark, A Woman of Color*, 1 Blackf. 122 (Ind. 1812). The most an aggrieved party (the cannery) can hope to get is damages, and those will likely be highly speculative and difficult to prove.

⁹⁸ See Bar-Gill and Ben-Shahar, "Law of Duress," 393-94.

⁹⁹ As Richard Posner put in a legal decision on this point, "[t]he sensible course would be to enforce contract modifications (at least if written) regardless of consideration and rely on the defense of duress to prevent abuse." *U.S. v. Stump Home Specialties Mfg.*, 905 F.2d 1117, 1122 (7th Cir. 1990).

¹⁰⁰ See Bar-Gill and Ben-Shahar, "Credible Coercion," 734-35 ("[T]he optimal policy is not to allow victims to sue for restitution of their robbed possessions, but rather to increase the likelihood of apprehending murderers and bringing them to justice as well as to increase the

sanction for murder.”). It should be noted that Bar-Gill and Ben-Shahar hedge on this point in that they state in passing that coerced contracts might be voided in cases where threatening parties deliberately make their threats more credible (e.g. by buying a gun), as they worry that not voiding these contracts would create incentives to invest in credibility enhancing means. *Ibid.*, 732-34. However, this qualification conflicts with their initial concern for protecting victims of credible coercion, since the victim of credible coercion will not care at all whether or not the threatening party *deliberately* invested in making his threat more credible. Consequently, it remains unclear where Bar-Gill and Ben-Shahar really stand with respect to enforcing contracts in the gun-man scenario. It does sound, however, as though they are prepared to bite the bullet and enforce such agreements when they are credible.

¹⁰¹ *Ibid.*, 734-35.

¹⁰² *Ibid.*, 722 (“A credible threat is one that the threatening party intends to carry out.”).

¹⁰³ *Ibid.*, (“If it is in the interest of the threatening party not to carry out the threatened outcome, his threat is not credible.”).

¹⁰⁴ *Ibid.*, 727 (applying the criterion to the highway gunman case).

¹⁰⁵ *Ibid.*, 737.

¹⁰⁶ See, e.g., Stephen J. Schulhofer, "Plea Bargaining as Disaster," *Yale Law Journal* 101 (1992): 1979 (“[P]lea bargaining seriously impairs the public interest in effective punishment of crime and in accurate separation of the guilty from the innocent.”). But see Robert E. Scott and William J. Stuntz, "Plea Bargaining as Contract," *Yale Law Journal* 108 (1992): 1918-24 (arguing that extant contract principles of coercion provide now reason to view plea bargaining as problematic).

¹⁰⁷ *Miller v. Fenton*, 474 U.S. 104, 116 (1985).

¹⁰⁸ *Lyons v. Okla.*, 322 U.S. 596, 601 (1944).

¹⁰⁹ *Garity v. N.J.*, 385 U.S. 493, 496 (1967).

¹¹⁰ *Frost Trucking Co. v. Railroad Comm'n*, 271 U.S. 583, 593 (1926).

¹¹¹ The same point can be made by analogy about the Exclusionary rule: punish police misconduct directly, but allow police to use the benefit of evidence obtained through such misconduct. *See Herring v. U.S.*, 129 S. Ct. 695, 699 (2009) (“[O]ur decisions establish an exclusionary rule that, when applicable, forbids the use of improperly obtained evidence at trial.”) One might also make the point about the entrapment defense. *See Jacobsen v. U.S.*, 503 U.S. 540, 549-50 (“Where the Government has induced an individual to break the law and the defense of entrapment is at issue, as it was in this case, the prosecution must prove beyond reasonable doubt that the defendant was disposed to commit the criminal act prior to first being approached by Government agents.”).