A PUZZLE ABOUT HOBBES ON SELF-DEFENSE

BY

CLAIRE FINKELSTEIN

Abstract: The crucial step in Hobbes' argument for the right of self-defense in civil society is the claim that the parties to the Covenant for the establishment of civil society do not abandon the right when they agree to lay down a portion of their natural rights. The reason they do not abandon this right is that it is not to their benefit to do so. But there seem to be many situations in which a person would benefit by abandoning the right of self-defense. I consider several interpretations that would rescue Hobbes' argument on this point. It turns out that the argument Hobbes needs to defend his appeal to benefit relies on a straightforward maximizing conception of rationality. I argue, however, that the account of rationality Hobbes presupposes in his answer to the infamous Foole is altogether different: it is a pragmatic theory of rationality instead. I argue that as between his argument for the inalienability of self-defense and his argument against free-riding on the Social Contract, Hobbes has no choice but to side with the latter. I also suggest, however, that Hobbes might have argued for the existence of a right of self-defense by a somewhat different route than the one he took. The reconstructed argument I propose depends upon a particular understanding of the right of self-defense in Hobbes’ system: It regards the right as the central political protection parties to the Covenant might regard themselves as having against the usurpations of an all-powerful sovereign.

1.

Are states required to grant their citizens a right to self-defense? Is the right to self-defense, in other words, a necessary feature of a legitimate political order? This question was once upon a time the subject of great interest in political philosophy, but it has fallen on hard times in recent years. Roughly since Locke, most writing on self-defense has focused instead on issues of relevance for moral philosophy. Much attention has
been lavished, for example, on whether the victim’s right to kill is compatible with the attacker’s presumptive right to life, whether the right can be exercised against innocent attackers, or whether the right can ever entitle a person to save her own life at the expense of an innocent bystander. But the political question might be thought prior to the various questions about the size and shape of the moral right. For the moral right is of little interest if it cannot be shown to supply a compelling reason for the State to grant its citizens a political right to the same. The question, then, is whether we can account for an obligation on the part of the State to respect whatever moral right to self-defense citizens turn out to have.

I trace the focus on the moral dimension of self-defense to Locke because he is the first simply to assert the relevance of the moral for the political right, without thinking it necessary to argue the point. Locke rests the civil right on two separate natural grounds. First, he argues, a person who attacks another, threatens to attack him, or even robs or threatens to rob him “pus him in a state of war with him against whom he has declared such an intention . . . ,” and Locke suggests that “I should have a right to destroy that which threatens me with destruction.” Second, Locke claims, there is a natural right “to judge of, and punish the breaches of” natural law in others. Where there is no common judge to resolve disputes, men revert to their natural right to punish the breaches of natural law against them. By ignoring the need to justify the move from these two natural entitlements to a civil right to self-defense, Locke set a precedent for overlooking a crucial question for any naturalistic account of the legal right to self-defense, namely the question of why the natural right to self-defense should carry over to civil society.

Earlier writings do not neglect the move from the moral to the legal right in the way that Locke does. Hobbes in particular devotes much attention to the question. Man in the state of nature, he thinks, has a right to everything he believes will conduce to his preservation, even “to one another’s body.” Such extensive possession of right, however, leads to a state of war. Thus man sees it as in his interest to lay down a portion of his rights, in order to leave his natural condition in favor of civil society. The right to self-defense, then, persists in civil society because it is not one of the rights that individuals must lay down as part of the original Covenant for the establishment of civil society. Not only is it not necessary to lay down this right, but Hobbes seems to think it could never be in a person’s interest to lay it down for any other reason. In short, Hobbes thinks the retention of the right to self-defense is rationally required, against the background of a view of human beings as rational maximizers.

Now there are many possible accounts of the relevance of moral for legal rights, and hence many possible accounts of the relation between the moral and the legal right to self-defense. My interest, however, is in
the contractarian account of this relation. And for this purpose, it is important to consider the merits of Hobbes’ argument carefully, for it is one of which any contractarian might avail himself. Indeed, Hobbes’ solution may well be the only one it is open to a contractarian to adopt. For if natural rights must be laid down voluntarily, as contractarians suppose, and if the civil right to self-defense is to stem from a natural right, the explanation for the civil right to self-defense must consist in an explanation of why it cannot be voluntarily abandoned. Put otherwise, since individuals must consent to the conditions under which they are governed, and since they give up no natural right they see as conducive to their protection and well-being to retain, a contractarian must appeal to some interest agents have in retaining the natural right if he is to explain the persistence of the right to self-defense in civil society. Thus despite the differences between Hobbes and Locke on many of the details of a contractarian political philosophy, one might suppose that both will require an argument for the inalienability of self-defense of roughly the sort Hobbes suggests.

My central claim in what follows is that in order to defend his argument for the inalienability of the right to self-defense, Hobbes would have to subscribe to a theory of rationality that we have reason to believe he did not hold and could not have held without substantially revising his political philosophy. Specifically, Hobbes would have to subscribe to the standard view of rationality as welfare maximization in order for his account of self-defense to go through. His answer to the important problem of freeriders to the social contract, however, depends on a different account of rationality—what is sometimes called the “pragmatic” account. I shall argue that although Hobbes’ argument for the inalienability of the right to self-defense fails, there is another version of the argument that works, one specific to the social contract itself. On this version of the argument, Hobbes must be able to explain why the parties to the Covenant would not see it as in their interest to retain various offensive rights they possess, even though they would think it in their interest to retain defensive rights. I shall suggest that the special importance of defensive rights lies in the fact that they provide citizens with a right of revolution citizens see it as in their interest to retain.

2.

The fullest presentation Hobbes offers of his argument for the retention of the right to self-defense appears in Leviathan Chapter XIV, paragraph 8:

Whensoever a man transferreth his right or renounceeth it, it is either in consideration of some right reciprocally transferred to himself or for some other good he hopeth for thereby.
A PUZZLE ABOUT HOBBES ON SELF-DEFENSE

For it is a voluntary act, and of the voluntary acts of every man the object is some good to himself. And therefore there be some rights which no man can be understood by any words or other signs to have abandoned or transferred. As, first, a man cannot lay down the right of resisting them that assault him by force, to take away his life, because he cannot be understood to aim thereby at any good to himself.

We might re-order the central premises of Hobbes' argument, for Hobbes appears to begin with an intermediate conclusion. Let us regard the argument as the following:

(i) A transfer of right requires a voluntary act.\(^9\)
(ii) The object of every voluntary act is some good to the agent who performs it.\(^10\)
(iii) No agent could think it good for himself to transfer his right to resist attackers.

Hobbes' conclusion is thus:

(iv) No man can be understood as transferring away his right to resist attackers.

We have not quite arrived, however, at Hobbes' conclusion for the necessity of a right to self-defense in civil society. In order to reach that conclusion, we need only help ourselves to a premise stating Hobbes' understanding of obligation. Since human beings start with unlimited liberties, it can only be impermissible for them to do something if they have voluntarily obligated themselves not to do it. Let us then add the following premise to the above argument:

(v) Voluntary transfer or abandonment of right is the only source of obligation in civil society.

It now follows that:

(vi) No man can be obligated not to resist attackers in civil society.\(^11\)

I shall focus on the third premise of the above argument, for I believe it is this premise that makes Hobbes' argument problematic. That is, I think his other premises can ultimately be defended, but the third premise cannot be. The problem with this premise is that it conflicts with Hobbes' most basic commitments about the nature of rationality. The question we will explore in connection with Premise Three thus strikes at the heart of the Hobbesian endeavor. Moreover, if I am correct that other contrarians will require something along the lines of Hobbes' argument for

© 2007 University of Southern California and Blackwell Publishers Ltd.
the inalienability of self-defense, considering Hobbes' difficulty with this premise may reveal a general obstacle to the contractarian account of self-defense. Before turning to the third premise, however, let me briefly consider the first two premises, since some may be inclined to reject Hobbes' argument for reasons connected with one of these premises instead.

The first premise seems sensible enough, although one would have liked a bit more argument in its support than Hobbes offers. In particular, one would have liked some sense of what Hobbes means by "voluntary act." Unfortunately Hobbes' definition of the term does not shed light on its use in this context. In Chapter Six, he tells us that a voluntary act is an act "which proceedeth from the will," and he explains that an action that stems only from inclination is not from the will and so is not voluntary. Beyond the latter restriction, his definition is intended to be quite open. Thus he explains that actions that proceed from "covetousness, ambition, lust or other appetites," as well as those that proceed from "aversion or fear of those consequences that follow" count as voluntary. An action need not be rational in order to be voluntary. We have somewhat more guidance from his remarks about voluntariness in contracts. Hobbes considers the case in which you make a contract with a robber to pay him money later if he will release his hold on you now. The question is whether you are bound to pay once he releases you. Standard contract doctrine would say the contract is not binding, on the grounds that it is coerced. But Hobbes disagrees, saying that the agreement with the robber is voluntary. How can an agreement made under threat of force be voluntary? Hobbes explains that it is voluntary because there is benefit on both sides. I will have more to say about this example later. For the moment, we need only take from it that when Hobbes speaks of a voluntary act, he means an act the agent could be said to choose as beneficial to himself. Understanding the notion of benefit rather broadly in this context, the first premise of the argument for self-defense can be understood simply as the claim that no one would transfer away a right without having some end in view to which he intends his act to contribute. By this, I take it Hobbes just means that no one could transfer away a right inadvertently.

The second premise, however, is likely to be more controversial. For the suggestion that a voluntary act is one the agent performs to achieve some good for himself would rule out altruistic acts as well as acts done for love of country or for aesthetic reasons. The premise as Hobbes states it thus appears to be an unvarnished appeal to psychological egoism, which many commentators find both unconvincing in its own right and inconsistent with Hobbes' larger purposes. Moreover, if Hobbes' argument for the right to self-defense in civil society depends on an egoistic account of motivation, other contractarians who do not share that assumption could not make use of it.
I do not, however, think that Hobbes' argument for the inalienability of the right to self-defense *depends* on psychological egoism. Thus, rather than join the debate over whether Hobbes is committed to psychological egoism, I suggest we avoid the difficulty by modifying his argument in a way that obviates this assumption. For the argument to go through, Hobbes need not have assumed that voluntary acts presuppose some self-interested good to the agent who performs them. He need only have assumed a weaker thesis, namely that an agent who performs a voluntary act hopes to attain *something he regards as a benefit*. We can allow Hobbes this claim while distancing ourselves from egoism if we say that the notion of benefit need not be restricted to enhancements of one's own well-being. A person, for example, might treat as a benefit the fact that someone she loves is bettered, without abandoning the basic picture of rationality as one of rational maximization. While not entirely uncontroversial in this form, the second premise would now have much wider appeal. So let us replace premise (ii) with the following:

(ii′): The object of every voluntary act is something the agent who performs it regards as a benefit.17

If we replace premise (ii) with (ii′), however, we will have a problem with the third premise. For as currently stated, it fails to connect with the revised argument, since it would not impair the voluntariness of a transfer of the right to self-defense that it was not egoistically motivated. So we should also re-write premise (iii), as follows:

(iii′) Transferring the right to resist attackers does not aim at anything the transferee would regard as a benefit,

where the notion of "benefit," once again, need not be interpreted egoistically. Let us then enhance the plausibility of Hobbes' overall argument by considering it in the following revised form:

(i) A transfer of right requires a voluntary act.
(ii′) The object of every voluntary act is something the agent who performs it would regard as a benefit.
(iii′) No agent could regard transferring the right to resist attackers as a benefit.
(iv) No man can be understood as transferring away his right to resist attackers voluntarily.
(v) A voluntary transfer of right is the only source of obligation in civil society,

with the conclusion that:

© 2001 University of Southern California and Blackwell Publishers Ltd.
(vi) No man can be obligated not to resist attackers in civil society.

These revisions are sufficient to make the first two premises reasonably palatable. It is, admittedly, another matter whether Hobbes would subscribe to the argument in its revised form. But this is not relevant for our purposes. The point of the above discussion is simply to make it reasonable to focus on the third premise, and to forestall rejection of the argument on the grounds of one of the other premises instead. Having satisfied ourselves that the first two premises are defensible under at least some interpretation, let us now turn to consider the crucial third premise of the argument in detail.

3.

The general claim represented in the third premise is quite a strong one. Hobbes is not simply saying that no one would likely agree to abandon the right to resist attackers voluntarily. He evidently means something more than this. For he thinks that even if someone did abandon the right by explicit words or deeds, believing he would benefit by doing so, we cannot understand him as having transferred the right, because it could never really be to his benefit to do so. Hobbes is not here making an empirical prediction about how agents actually behave. He is making a conceptual claim about what rationality requires. He is saying that we cannot both understand a person as voluntarily giving up his right to self-defense and continue to regard him as a rational agent. I shall call this the argument from benefit.

Put this way, however, the premise is surely false. For one can readily imagine situations in which a person might seek to benefit by abandoning his right to resist attackers. Suppose you have a knife to my throat, and you threaten to kill me now unless I agree to give up my right to self-defense for all future occasions. It would clearly be to my benefit to accept the terms of this agreement for the sake of avoiding certain death now. At least, as Hobbes would say, there is time of life gained. So the question is why, in the face of obvious examples to the contrary, Hobbes thinks it cannot be to a person's benefit to transfer away the right to self-defense.

The argument from benefit becomes even more puzzling when we consider its extension to "wounds, and chains, and imprisonment," which Hobbes goes on to make after the portion of Chapter XIV, paragraph 8, quoted above. He says that "there is no benefit consequent to such patience," parallel to the argument he made about conveying away the right to resist attackers.\textsuperscript{18} But surely it would sometimes be beneficial for me to
lay down the right to resist him who wounds me, puts me in chains or imprisons me. Indeed, it might not take very much to induce me to abandon this right. There is probably some large sum of money someone could offer me that would make me regard abandoning it as worthwhile, especially if the imprisonment were of limited duration. At least at first blush, then, Hobbes’ argument that human beings retain the right to self-defense in civil society because it could not be to their benefit to transfer it away fails, and the argument seems all the more erroneous when applied to the right to resist wounding and physical restraint.

The difficulty with Hobbes’ argument from benefit is so obvious that we might doubt that Hobbes meant us to take it at face value. Indeed, there appear to be at least two alternate interpretations of the third premise we should consider, either of which would make Hobbes’ claim look more plausible. I shall call them the argument from involuntariness and the argument from incapacity, respectively. While both arguments may seem initially quite sensible, I shall suggest that Hobbes could not in fact have intended either.

First the argument from involuntariness. Someone might argue that it could only be in a person’s interest to abandon his right to self-defense if he were under immediate threat to his life, but that any agreement made under threat to one’s life must be considered void on grounds of lack of voluntariness. In other words, any rational agreement to abandon the right to self-defense would necessarily be coercive. It is important to note that the argument for the invalidity of such contracts is normative, not psychological. There are certain background conditions we require in order to regard a person as having acted of his own free will, and the notion of voluntariness might be thought of as capturing these conditions. When we say, therefore, that a contract signed under duress is “involuntary,” we do not mean literally that the actor did not choose to enter the contract. Rather, we mean that the background against which he chose was one that makes his choice fail to reflect his agency. Perhaps, then, when Hobbes says that any agreement to abandon one’s right to self-defense must be involuntary he is making a normative claim about the limitations on enforceable agreements.

Now there are a number of reasons to reject this interpretation, some specific to Hobbes, others of a more general nature. Let me just mention one reason of each sort briefly. First, Hobbes is quite clear in maintaining that covenants entered into out of fear are not invalid. He makes the point most explicitly in the example of the agreement with the robber we considered earlier:

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.\footnote{20}

Shortly thereafter, Hobbes extends the principle to civil society: "And even in 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."\footnote{21} Hobbes makes the point even more forcefully in De Cive. 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 100L. 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 proceeded 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... \footnote{22} It holds universally true, that promises do oblige when there is some benefit received...

As we saw earlier, Hobbes rejects the suggestion that the agreement with the robber is involuntary, because what is crucial for him is the fact that it is mutually beneficial. Mutual benefit provides the explanation both of why it is voluntarily entered into, and of why it is binding. It is voluntary, because the test of voluntariness in actions is expectation of benefit. And it is binding simply because it is a voluntary renunciation of a liberty—the liberty not to pay the robber. For Hobbes, in other words, voluntary renunciation is not only necessary to create an obligation not to exercise the right; it is also sufficient. Since there are clear cases in which it would be a benefit to the transferor to relinquish the right to self-defense, and since Hobbes accepts that a person can act from benefit in agreeing to a coercive contract, it is not open to Hobbes to defend the third premise with the argument from involuntariness.

Quite apart from what it is open to Hobbes to argue, there seem to me to be independent reasons to reject the argument from involuntariness. I suspect, in other words, that Hobbes was right to reject the idea that actions performed under duress are involuntary. First, a coercer depends on his victim’s ability to respond to his demands rationally for his efforts at coercion to be effective. He relies on his victim to engage in a rational balancing of evils, and it would be odd to say that an action that is rational is nevertheless not voluntary. Second, it is hard to distinguish these kinds of cases from those in which one’s options are simply unattractive. If I need, desperately need, to buy a loaf of bread, and there is only one kind of bread in the store, we would not normally say I bought that kind under duress. It would be worse, still, to say I bought it involuntarily. But what possible difference could it make that in one case a person has
a gun to my head, and in the other, I have compelling reasons of a
different sort to favor a particular course of action? I have discussed this
issue elsewhere, however, and so I shall not press the point further here.23

4.

Let us now turn to the second alternative defense of Hobbes' third
premise—what I have called the argument from incapacity. It might seem
that Hobbes is really suggesting that it would not be psychologically
possible to comply with an agreement to abandon the right to self-defense,
however much such an agreement would be to one's benefit. The imposi-
sibility would stem from the fact that the instinct of self-preservation is
too strong for human beings to make good on any agreement that would
require them not to defend their lives when attacked. And if this is true,
then I could not honestly undertake an obligation not to defend myself
when attacked. For I cannot sincerely promise to do something I know
I lack the capacity to do. If someone offered me eternal happiness in
exchange for my promise to play the piano portion of Rachmaninov's
Second Piano Concerto this evening, it would very much be in my inter-
est to do so. But its being in my interest does not make it possible for me
to do so, and thus any promise I might make could only be insincere,
since I know I lack the capacity to make good on such a promise. The
claim that no one could agree to abandon the right to self-defense, under
this interpretation, would thus be a claim about capacity. It would be the
claim that no matter how hard I tried not to defend myself when attacked.
I would be incapable of not striking back. Any promise I might make not
to defend myself is thus as trustworthy as a promise to play the piano
portion of the Rachmaninov tonight.

There is somewhat more textual support for this interpretation of
Hobbes' third premise than there is for the argument from involuntariness.
We have this from Chapter XIV of Leviathan:

For though a man may covenant thus unless I do so, or so, kill me, he cannot covenant thus
unless I do so, or so, I will not resist you, when you come to kill me. For man by nature
chooseth the lesser evil, which is danger of death in resisting, rather than the greater, which
is certain and present death in not resisting.24

This passage seems quite different from the argument Hobbes gives us
earlier in the chapter (paragraph 8). For here Hobbes seems to acknowledg-
that an agreement that amounts to giving up one's life at a later
time can be worthwhile if it is done for some greater good now—that is,
he seems to admit the defectiveness of the argument from benefit. Instead,
he suggests that advantageous though it might be, I cannot contract

© 2001 University of Southern California and Blackwell Publishers Ltd.
away my right to resist you when you come to take my life, since I will be \textit{incapable} of simply standing there and accepting your blows.

But why should we think it impossible to stand there and accept another’s blows? For no matter how unpleasant, it is presumably not like playing a piece on the piano one lacks the ability to play, or like any other physically impossible act, such as unaided flight. On the face of it, the claim that it is simply impossible not to defend oneself when attacked seems implausible. The difficulty associated with failing to defend oneself seems rather that it would not be to one’s \textit{benefit} not to defend oneself, and this has more to do with rationality than with capacity.

To the extent we can find any defense in Hobbes of this odd claim about capacity, it is his suggestion that man always chooses the lesser over the greater evil, and that fighting is a lesser evil than letting oneself be killed. But this argument does not seem to vindicate the claim about capacity. For abandoning the right to self-defense may itself be the lesser evil, even factoring in the costs of compliance. Most importantly, the appeal to lesser evils seems once again to invoke the notion of benefit, rather than psychological impossibility. For the only kind of psychological “impossibility” there could be is one that comes from signing up for something that is not to one’s benefit. This would make the claim about incapacity one about rationality once again, and so we would not have sketched an alternative to the interpretation from benefit.

What this seems to show is that where psychological, rather than physical, impossibility is concerned, the argument should run \textit{from benefit to capacity}, rather than the other way around. That is, if we start from the premise that human beings are rational agents, we should attribute to them the capacity to do that which it is to their benefit to do. Of course this claim must operate within certain physical and intellectual limits: we cannot attribute the capacity to agents to do everything they would benefit from doing. But the instances in which we cannot make this move, like playing Rachmaninov’s Second, are distinguishable on the grounds that they require skills of a certain sort, skills that no amount of benefit can provide. We should not associate the difficulty of standing there and accepting physical blows with an absence of skill.

Having dispensed with both the argument from involuntariness and the argument from incapacity, let us turn to what I take to be Hobbes’ real argument for the infalienceability of the right to self-defense—the claim that no person could rationally regard herself as benefited by abandoning the right to self-defense. This, I think, is the right way to understand Hobbes’ claim in Premise Three. Quite apart from the difficulties with any other interpretation of his claim, the argument from benefit squares with the fact that Hobbes seems to be presenting us with a theory of political arrangements based on rational agency. That is, if we understand Hobbes’ project as an investigation into human rationality, the appeal to
benefit makes Hobbes’ argument for the inalienability of the right to self-defense consistent with his larger project.

5.

The most natural defense of Hobbes’ argument from benefit is that it cannot be to anyone’s benefit to agree to give up the right to self-defense, because it cannot be to anyone’s benefit actually to give up the right. And it cannot be to anyone’s benefit actually to give up the right, because it cannot be to anyone’s benefit to be without the right. For if I lack the right to defend myself, then I cannot defend myself if attacked. But if attacked, it would never be in my interest not to defend myself. It would follow, then, that I cannot benefit by giving up the right to self-defense, and so any agreement to do so is void.

Two objections might be raised at this point. First, could it not be in my interest to agree to give up the right to self-defense, even if it were not actually in my interest to give it up? For if I could convince you that I mean to give up the right, even if I do not, then you would release your hold on me now. So it does not look as though the argument from benefit itself would rule out insincere promises. On what grounds, then, is Hobbes ruling them out?

Second, why should we think that giving up the right to self-defense entails actually not defending oneself if attacked? Does it follow from the fact that I give up a right to do something that I cannot do it? Can I not simply act without right? So why not allow that a person could sincerely agree to abandon the right to self-defense, and sincerely abandon it, but still defend herself at some future point if attacked?

With regard to insincere promises, it would of course be in my interest to agree to give up the right if I could do so without actually giving it up. But there is reason to think insincere promises are infeasible. Since we are exploring a theory of ideal rationality, we must assume that both parties to any such agreement are rational, and that each knows the other is rational. Then if I know it is not in my interest to give up the right to self-defense, you will also know it is not in my interest to give it up. It follows that you will also know that any promise I make to give up the right is insincere. The only way you will believe me when I promise to abandon the right to self-defense is if you believe it is in my interest to abandon it, and the only way you will believe that is if it is in fact in my interest to do so. For these purposes, then, we can assume that only sincere promises will be binding.25

With regard to acting without right, there are several reasons to reject this objection. First, recall that we are considering Hobbes’ argument for the right to self-defense in civil society. If we allow that people can act
without right as part of the argument for what rights it is rational for them to retain, we will be unable to consider the effects of any abandonment of right. In order to consider the merits of retaining particular rights, then, we must assume that people will not act without right in civil society. We can then treat separately the problem of those who act without right, namely defectors from the agreement to abandon rights. This in fact is precisely what Hobbes does.

Second, on a Hobbesian understanding of rights, it is probably not correct to think that abandoning a right is compatible with harboring an intention to engage in the acts the right licenses. For Hobbes thinks of a right as a liberty to do something without impediment. When a person abandons a right, she binds her will such that she deprives herself of the power to act in the way the right describes. Hobbes makes this apparent in his explanation of why the original grant of power to the sovereign cannot simply be revoked upon the mere displeasure of the citizens.\textsuperscript{26}

Third, and most importantly, even if we were to allow that a person could abandon a right to do something in Hobbes’ system and still intend to do it, the same difficulty we saw above would emerge: If it were rational for me to agree to abandon the right while intending to act as though I had retained it, you would know it was rational for me to do so. If I am to convince you not to kill me when you have a knife to my throat, then, you must believe not only that I am sincere in abandoning the right to self-defense, but that my abandonment is incompatible with my intending to defend myself if attacked. You must believe, in short, that my renunciation of right constitutes a sincere agreement on my part not to defend myself when you come to attack me. If I cannot convince you of this, I cannot save my life. And if this is true, then once again it cannot be in my interest to abandon the right to self-defense, since that would be to commit myself to act in a way it would not be in my interest to act.

We may assume, then, that if it is not in my interest not to defend myself if attacked, it is not in my interest to abandon the right to defend myself if attacked. And if it is not in my interest to abandon the right, then it is not in my interest to agree to abandon the right. Of course this is all rather tragic from my point of view. For all things considered, I would trade my future right of self-defense for my life now. While I would most like to convince you to release me without my having actually to give up the right, I would still regard a deal under which I gave up the right in exchange for my life now as advantageous. That is, while I could not regard giving up the right to self-defense as advantageous considered in itself, I regard the package—abandon the right to self-defense in exchange for freedom now—as advantageous, since it is the best of my available alternatives. The question is how to make it rational for me to abide by such an agreement. For that is what I need to convince you that
I would indeed make good on a promise to give up my right to self-defense. And if I cannot convince you of my commitment to abide by such an agreement, I cannot induce you to release me now.

My difficulty stems from the fact that no matter how advantageous the terms of the agreement are to me ex ante, I will have reason to reconsider my end of the bargain once you have released me. And if I reconsider, I will see that I have no reason to stand there and accept your blows without defending myself, even though it was advantageous for me to agree (sincerely) to do so at the time we struck our deal. The core of my problem, then, is that although the agreement is advantageous to me by its terms ex ante, this does not suffice to make it rational for me to execute those terms ex post. And without its being rational for me to make good on such terms at the time of performance, I cannot convince you that I will adhere to our agreement.

Matters would naturally be different if I could precommit to not defending myself when you came to attack me later. If I could pay someone to force me not to defend myself, for example, I could reassure you of the sincerity of my promise not to defend myself at the time of a future attack. Precommitment would foreclose reconsideration of the terms of our agreement, thus rendering the agreement advantageous for both parties. Matters would also be different if I hoped to benefit from future relations with you, for you would be reassured by the incentive I had to abide by our agreement. While not actually barred from reconsideration, I would have no incentive to reconsider, and you could expect me to perform when the time came. But assuming no precommitment strategy is available, and given that there will be no future course of dealings in our case (since I will likely be dead in the near future in any event), it looks as though I cannot sincerely enter into an agreement with you that will allow me to live now in exchange for not defending myself later.

The problem with abandoning the right to self-defense under the circumstances imagined is that presented by a familiar rational choice problem, the toxin puzzle. A billionaire, who is a very good judge of intentions, offers to put a million dollars in your bank account if you intend by midnight tonight to drink a certain toxin tomorrow. The toxin will make you quite ill, but it will not kill you or have any long-lasting effects. The problem seems easy to solve: you should commit to drink the toxin and you will be rich. The only difficulty is that the billionaire has cleverly stipulated that you need not actually drink it in order to have the money deposited in your bank account. You need only form the intention to drink it. And when you think about what that means, you realize you are in a bind. For come tomorrow, the money is either in the bank or it is not. Actually drinking the toxin thus cannot make you rich and can only make you sick. So it cannot be rational for you to drink the toxin tomorrow. And since you know that now, you cannot form the intention
to drink it, since you cannot form an intention to do something you know you will have no reason to do. Thus although you would trade sickness for a million bucks, it looks as though you have no way of effectuating such a trade, advantageous though it might be. 28

In the case we are considering, I am in the position of the person offered the billionaire’s deal when, with knife to my throat, I agree to abandon my right to self-defense in exchange for my life now. Since it cannot be rational for me to make good on such an agreement at the time of performance, I cannot convince you to let me survive. For sincerely agreeing to abandon the right to self-defense requires me also to form the intention to refrain from defending myself when the time comes. But how can I form the intention to do something I know it will be irrational for me actually to do? So ironically I must die now, even though you would be prepared to offer me a deal under which I might live, and which would be to my benefit to accept, if I could assure you I would carry through. Since I have no way of providing you with the assurance you need, I will not be able to convince you to refrain from killing me now.

It looks, then, as though we have found a way to support Hobbes’ assertion of Premise Three. Given that it is not rational for an agent to refrain from defending herself when attacked, it cannot be rational for her to abandon the right to defend herself now. And if it is not rational for an agent to give up the right to defend herself, it cannot be rational for her to agree to give up the right. Thus despite the fact that one might like to be able to commit oneself to refraining from defending oneself if attacked, Hobbes seems right to consider any agreement to abandon the right to self-defense void, since we cannot suppose that a rational agent would ever actually make good on such an agreement.

6.

But let us consider the matter a bit more carefully, for we may have given Hobbes his premise too quickly. After all, not everyone agrees that it is impossible to win the mad billionaire’s gambit. Some philosophers think you should simply decide you will drink the toxin come what may, and then actually drink it. After all, you would precommit to drink it if you could, for example by paying someone to force you to drink it when the time came. Would it not, then, be more advantageous simply to form the intention to drink by planning to drink, and then to refuse to reconsider that intention? That way, you could reap the benefits of precommitment without incurring its costs. A rational agent should thus regard the plan that involves sincerely committing to drink the toxin and then not drinking as infeasible. She should see the second-best plan of committing to drink

© 2001 University of Southern California and Blackwell Publishers Ltd.
and actually drinking as the best she can do. A person who reasoned in this way would have adopted what we might call a pragmatic approach to rationality—an approach that assesses the rationality of available options strictly according to the welfare of the agent under each option. It imposes no criteria of rational assessment that is independent of the outcome of each possible choice an agent might make.

Let us call a theory of rationality that endorses the pragmatic approach the “pragmatic theory of rationality.” The pragmatic theory of rationality suggests evaluating the merits of actions in terms of whatever will allow one to fare best. While in many cases, the pragmatic theory will recommend the same course of action as a straightforward welfare-maximizing account, sometimes the two will diverge. In particular, the two will diverge when one can do better—make more money, for example—if one reasons about actions by placing them in the context of available plans. For in that case, the pragmatic theory assesses actions relative to the plans of which they are a part, whereas the welfare-maximizing approach evaluates actions in isolation, irrespective of their position relative to plans. In the toxin puzzle, the pragmatic approach leads to the conclusion that you should actually drink the toxin, since that is a required part of the only course of action you can adopt that will actually allow you to get the money. On a welfare-maximizing conception, by contrast, drinking the toxin would be irrational, since the act must be assessed in isolation, irrespective of its position in the overall plan of which it is a part.

If Hobbes were to avail himself of the argument I offered on his behalf in support of the third premise, he would have to reject the pragmatic theory of rationality. For he would have to endorse the rationality of one course of action, when another course of action, considered as a whole, would be more advantageous. In particular, he would have to endorse the rationality of retaining the right to self-defense, even when an agent could expect to do better under a plan that involved abandoning it. The defensibility of Hobbes’ third premise goes to the heart of the contractualist endeavor, then, because it requires us to articulate the contractualist’s underlying conception of rationality. In particular, it requires us to determine whether the contractualist is a pragmatist about rationality. If he is, the argument for the inalienability of the right to self-defense will not succeed.

On the face of it, the pragmatic conception of rationality should appeal to a Hobbesian. For a Hobbesian ought to agree that it is rational to do whatever would most conduce to one’s preservation, thus strongly suggesting an outcome-oriented account of rationality. The point has been little noticed, for Hobbes is thought to hold the standard welfare-maximizing account of rationality. But Hobbes is emphatic in seeing rational agency as deliberation in the service of one’s welfare. When welfare maximization
and outcome rationality conflict, the Hobbesian should presumably side with the latter, since otherwise he must endorse the rationality of plans and courses of action that leave an agent worse off. But then, as I have claimed, Hobbes cannot defend the third premise in his argument for the existence of a right to self-defense. My claim is that Hobbes must choose between the pragmatic account of rationality and his argument for the existence of a right to self-defense.

Framed in this way, the choice may seem an easy one. The pragmatic theory of rationality is highly tendentious. Saddling Hobbes with an unlikely account of rationality and then showing that his commitment to it creates problems for another part of his political philosophy hardly seems a serious challenge. Presumably he should fix the problem by abandoning the unlikely account of rationality in favor of the standard, welfare-maximizing model. Nevertheless, I shall suggest that Hobbes cannot easily relinquish his commitment to the pragmatic account of rationality. My claim, in brief, is that he cannot reject it without substantially impairing his ability to solve the problem of freeriders to the social Covenant—the challenge of the infamous Fool. And if I am right, then Hobbes' difficulty is a serious one. For the choice he then faces is between his solution to the problem of freeriders and his account of self-defense. Both are crucial parts of his political philosophy, and neither, it would seem, can be easily abandoned.

7.

The Fool's challenge is an attack on the third law of nature, which exhorts men to keep their covenants. The Fool says it is not rational to keep covenants in any case in which one can expect to do better by breaking them. Hobbes has him say:

[Every man's conservation and contentment being committed to his own care, there could be no reason why every man might not do what he thought conduced thereunto, and therefore also to make or not make, keep or not keep covenants was not against reason, when it conduced to one's benefit.]

This argument presents an important difficulty for Hobbes' account. For given that the foundation of civil society is contractual, the Fool would have men defect from the bonds of civil society whenever it was advantageous for them to do so. Moreover, it looks as though it would always be advantageous to defect, since once one person had laid down his rights, it would always be in another's interest to attempt to dominate him if he thought it likely he would succeed. The problem, then, is that no matter how great the benefits of agreeing to abandon one's natural
rights, it looks as though it could never be rational to abide by such an agreement.

It is a frequent refrain among commentators that Hobbes' answer to the Fool is inadequate. David Gauthier, for example, says that Hobbes' response "tends to miss the point of the objection." Jean Hampton describes the reply as "remarkable," on the grounds that it "directly contradicts the position taken in the [previous] chapters." The reply, however, seems to me to be more substantial and better constructed than Hobbes is given credit for. To see this, his answer must be read through the lens of the pragmatic account of rationality.

The salient part of Hobbes' immediate reply to the Fool is as follows:

[First, that when a man doth a thing which, notwithstanding anything can be foreseen and reckoned on, tendeth to his own destruction . . . , yet such events do not make it reasonably or wisely done. Secondly, that in a condition of war wherein every man to every man . . . is an enemy, there is no man can hope by his own strength or wit to defend himself from destruction without the help of confederates . . . ; and therefore, he which declares he thinks it reason to deceive those that help him can in reason expect no other means of safety than what can be had from his own single power. 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. . . .]

Hobbes is suggesting that defecting from an agreement which it was to one's benefit to enter does not conduce to one's preservation, but rather promotes one's destruction. Defection tends to one's destruction, moreover, even if one ends up gaining the benefits from defection for which one had hoped. How can this be? Suppose a person can be assured that he will never be detected and can successfully free-ride on the efforts of others if he defects from a cooperative arrangement with them. What could be irrational about defection under these circumstances? Hobbes clearly thinks defection would be irrational, and his explanation appeals to the pragmatic theory of rationality.

Hobbes' first move is to remind the Fool that he is talking about a law of nature. That is, he emphasizes that what is at stake is a general precept of rationality. His basic argument is that there can be no general rule of rationality that promotes breaking covenants when it suits one to do so, for the reason that a person who would violate an agreement is a person with whom it would be irrational for others to covenant. For "he cannot be received into any society that unite themselves for peace and defence but by the error of them that receive him. . . ." Hobbes suggests a different precept of rationality instead, namely that if it was rational for an agent to enter into an agreement in the first place, it is rational for him to keep the agreement, since otherwise he can only count on being included

© 2001 University of Southern California and Blackwell Publishers Ltd.
in agreements generally on the supposition that others are mistaken about whether he is a deceptor or an agreement-keeper. We cannot assume that others are irrational, however.\textsuperscript{24} Hobbes thus argues for the rationality of particular acts of sub-optimal covenant-keeping by appealing to the rationality of a general rule in favor of keeping covenants. In this way, his defense of the third law of nature presupposes the pragmatic approach to rationality.

There are, admittedly, parts of Hobbes’ response that one might marshal in defense of a standard welfare-maximizing interpretation. For example, Hobbes’ suggestion that backing out of the conditions of civil society “tendeth to [a man’s] destruction” may sound like an appeal to expected benefit. And there is the empirical ring of Hobbes’ emphasis on the person who declares he is entitled to break his covenant, as opposed to the person who breaks a covenant silently and is never discovered. Presumably it is because of passages like these that even a committed proponent of the pragmatic theory of rationality like Gauthier is led to conclude that “Hobbes insists that covenants are binding only because it is in one’s interest, based on reasonable expectations, to keep them.”\textsuperscript{35} But I do not think even these passages compel the expected benefit reading of Hobbes’ answer to the Fool.

First, the reason an act of covenant breaking “tendeth to [one’s] own destruction” is arguably not that the expected benefits of covenant breaking outweigh the costs, but that it is a condition on the rationality of making covenants that covenants be kept, and the ability to enter into covenants tends to one’s benefit. The point is analytic, not empirical. The emphasis on declaring the entitlement to break covenants, as opposed to merely silently breaking them, is also susceptible to interpretation in pragmatic terms: If it were rational to break covenants, it would have to be rational to subscribe to a principle of covenant-violation. But no one could subscribe to such a principle and still be the sort of person with whom others could covenant. Thus while there are sentences in Hobbes’ response to the Fool that could be read in expected benefit terms, there are none that cannot also be interpreted through the lens of the pragmatic theory of rationality. And given that Hobbes’ response does seem to miss the point of the Fool’s objection when interpreted as a claim about expected benefit, as Gauthier says, there seems no reason to read it in this way.

There is further support for the pragmatic approach in Hobbes’ general remarks on contracts. Hobbes draws a distinction between two kinds of contracts: one that is valid in a state of nature and requires no sovereign for its enforcement, and another that is not valid in a state of nature and does require a sovereign for its enforcement. Hobbes scholars have paid little attention to the difference between the two, but both turn out to be crucial for Hobbes’ political theory. The former is necessary because
if a sovereign were required for the validity of every contract, then the
sovereign could not himself be authorized contractually. The latter is
necessary because if all contracts were available in a state of nature, it is
hard to see what purpose the sovereign would serve.\textsuperscript{36} The form
the distinction takes in Hobbes is that between contracts where one party is
to perform now and the other is to perform in the future, and contracts
where both parties are to perform in the future. "First-performer contracts"
are valid in a state of nature, Hobbes says, but "future-performance
contracts" depend for their validity on civil enforcement. Hobbes therefore
answers the Fool by conceding his objection to one kind of contract—
future-performance contracts—but pointing out that there is another kind
that is not subject to his critique—first-performer contracts. And he sugges-
tests that the kind of contract he requires for his political argument to
work is the first-performer contract, so the Fool's objection is not ulti-
mately damaging to his account.

Now the expected benefit reading of Hobbes' answer to the Fool looks
especially implausible when it is considered in conjunction with Hobbes'
distinction between these two kinds of contracts. For why would the two
differ if the rationality of keeping covenants were determined by expected
benefit? In particular, Hobbes is quite clear that he accepts the expected
benefit reading of contracts for future performance. Why would he bother
distinguishing first-performer contracts from future-performance contracts
if he intended expected benefit to determine the rationality of both?

But if not by expected benefit, how does Hobbes distinguish between
first-performer and future-performance contracts? Hobbes does not explain
the matter very thoroughly, but consider his account of why first-
performer contracts are binding:

[H]e that promiseth only (because he hath already received the benefit for which he promiseth)
is to be understood as if he intended the right should pass; for unless he had been content
to have his words so understood, the other would not have performed his part first.\textsuperscript{37}

And, he explains further, "[h]e that performeth first in the case of a
contract is said to MERIT that which he is to receive by the performance
of the other, and he hath it as due."\textsuperscript{38} Hobbes' thought, then, seems to be
that there is a natural entitlement to performance on agreements on which
one has already performed, an entitlement that depends only on the fact
of one's own good faith performance, rather than on the coercive power
of civil society to enforce the contract. First-performer contracts are thus
binding on something akin to a reliance theory. They are binding, once
one party has already performed, because the first-performer has relied
to his detriment on the promise of the other party. But why does reliance
on the part of the first-performer create an obligation for future perform-
ance on the part of the other party? Hobbes' answer should be familiar: it

© 2001 University of Southern California and Blackwell Publishers Ltd.
is rational to be a second performer on a first-performer contract, because the person who has already performed would have had to be irrational had he done so in the face of its being rational for you not to perform. And since we cannot assume the irrationality of any man, we cannot suppose that it would be irrational for you to keep up your end of the bargain.

We can now read Hobbes' answer to the Fool in light of his defense of being a second performer in contracts for mutual advantage where the other party has already performed. The original social Covenant is such a contract at the point at which Hobbes' Fool raises his objection. The second law of nature has already exhorted men to seek peace, and thus to lay down their rights when others are willing to do the same. The third law—that men keep their covenants—applies where rights have already been laid down, i.e. where performance on the part of others has already occurred. The question is whether you should keep up your end of the bargain by continuing to leave your rights in abeyance while others are vulnerable. Hobbes says it is rational to be a second performer in this case, for others have already relied on your promise of future performance by laying down their rights. The Fool is a fool because he fails to distinguish the future-performance contract from the first-performer contract, thinking that expected benefit provides the basis for compliance in both situations. He is a fool, in effect, because he fails to distinguish the situations to which the second and the third laws of nature apply. For he takes the injunction of simultaneous performance pertaining to entering into contracts contained in the second law of nature and applies it to the exhortation to keep covenants that have already been entered into in the third law.

There are at least two problems with Hobbes' argument, which I will address briefly, although I cannot do full justice to them here. First, why does the reliance argument not apply to covenants for future performance as well? That is, one might argue that since you would not have received the benefits of his cooperation if it were not rational for you to make good on your end of the bargain, failing to make good on a contract for future performance presupposes the other's irrationality, which presupposition cannot be assumed to be true. So if you are both rational and each has knowledge of each other's rationality, then the other party would not have contracted with you—even for future performance—were it rational for you not to perform. Since he did contract with you, it appears to follow that it would not be rational for you not to perform. From this one can conclude that it is rational to perform on any agreement, even when the other party has not yet performed—a conclusion Hobbes must reject, as I have argued above. The reliance argument, therefore, must be wrong, for it would ultimately defend the rationality of agreements that Hobbes thinks are void in a state of nature.
Second, if it is rational to be a second performer, must it not also be rational to be a first performer? For one party can apparently make it rational for the other to perform, just by performing first himself. But this implies that any contract for future performance can be converted into a first-performer contract: all that is necessary is for one of the parties to perform, thus obligating the other. If this is correct, then the distinction between first-performer and future-performance contracts falls apart, since it depends on a spurious distinction between the rationality of first and second performance. So how, exactly, does Hobbes distinguish the rationality of first as against second performance, in the face of an argument that the rationality of the latter entails that of the former?

These two problems arguably have a common solution, for Hobbes’ reasons for distrusting contracts for future performance are precisely the same as his reasons for thinking first performance irrational. Both the person who agrees to a contract for future performance and the person who agrees to act now for the sake of the other’s subsequent performance are uncertain of the gains from their own performance. There are any number of reasons why the other party might fail to perform: he might turn out to be irrational or his circumstances might have changed. The second performer, by contrast, is certain he is benefitting from the agreement, even counting the costs of his own performance. For he has already gained the benefits for which he entered into the agreement, since the other party has already performed. Performance on his part is not the least bit uncertain or risky, and thus cooperative behavior is not irrational under these circumstances. Indeed, this thought is in the second law of nature itself, in the admonishment to lay down rights only insofar as others are willing to do the same. Hobbes thus does seem to want to defend the counter-intuitive proposition that it is rational to perform second even though it is irrational either to perform first or to commit oneself or to perform where the other has not yet performed. This allows him to protect the distinction between first-performer and future-performance contracts.

A final problem with this argument, however, deserves mention. If it is irrational to be a first performer, does that not suggest there will never be a first-performer contract in a world of perfect rationality? Why doesn’t Hobbes regard this as invalidating his argument for the social Covenant, since, as we have seen, the Covenant is a kind of first-performer contract? I suspect the answer lies in the fact that the Covenant is a very special kind of first-performer contract. On the one hand, as Hobbes imagines it, there is initially a simultaneous laying down of rights. This by itself does not require any kind of contract, since it is all present performance, and there is no performance still outstanding once rights have been laid down. On the other hand, there is a more future-oriented agreement the parties make to leave their rights in abeyance for all time, and it is this part that
I suspect Hobbes thinks of as a first-performer contract. For each person has laid down his rights now on the condition that others abandon their rights and leave them in abeyance in the future. Each, therefore, is both a first and a second performer: each is a first performer by virtue of laying down his rights, and each is a second performer by leaving his rights in abeyance based on the other parties' laying down of right. While, as I have argued, Hobbes seems to think first performance is generally irrational, this instance of first performance is rational. For the laying down of right is done not as the first move in a first-performance contract, but as a non-contractual exchange with others that is beneficial to all participants.

The problem of the distinction between first-performer and future-performance contracts is a complex one, and my brief remarks on the subject do not do it justice. My hope, however, is that they have been adequate to sketch an interpretation of Hobbes' reply to the Fool that shows it to be a strong, a priori response, rather than the weak appeal to expected benefit it is normally taken to be. If my interpretation is correct, Hobbes' solution to the problem of freeriders to the social contract requires commitment to the pragmatic account of rationality. For Hobbes must be able to argue for the rationality of being a second performer, and this requires exactly the sort of reasoning from the merits of the plan, taken as a whole, that the pragmatic theory advances. Thus Hobbes could not easily dispense with the pragmatic account of rationality, for his treatment of the stability of the original Covenant depends upon it. In effect, all contracts would be like contracts for future performance in the absence of the pragmatic theory of rationality. It follows that Hobbes' argument for the inalienability of the right to self-defense fails, as I had initially feared it would.

8.

Perhaps, however, Hobbes does not need to show that it would always be irrational to give up the right to self-defense. For a more limited claim might suffice, namely that the parties to the social Covenant would not have regarded it as advantageous to give up the right. For this, Hobbes would require a positive political argument relating to the specific interests of parties to the Covenant.

Any such argument would face an important difficulty, stemming from the fact that the right to self-defense is not the only right of self-preservation man has in the state of nature. He has, in addition, an offensive right to attack others who are not presently attacking him, if his preservation, in his view, requires it. Indeed, the right is stronger than that. He may even attack someone who is not presently attacking him if his well-being
would merely be increased by the attack; his preservation need not depend on it. The question then arises why the parties to the Covenant would regard it as advantageous to retain the right to self-defense, without regarding it as similarly advantageous to retain a series of offensive rights. That is, the argument we reconstructed from Chapter XIV, paragraph 8, might be easily modified by making the following substitution for premise (iii):

premise (iii*): No agent could regard abandoning the right to use the bodies of any man as a benefit.

Hobbes' conclusion would then be:

premise (iv*): No agent can be understood to have abandoned his right to use the bodies of others.

And it would follow that:

premise (v*): No agent can be obligated to refrain from using the body of any other man in civil society,

which Hobbes must reject. If we cannot distinguish self-defense from other forms of self-preserving acts, the argument for the retention of the right to self-defense will threaten to cast us back into a state of insecurity and fear, and thus dissolve the bonds of civil society. Against the background of this problem, it is not surprising that Hobbes tries instead to establish a general conceptual argument pertaining to self-defensive rights across the board.

What we would require, then, to make the more specific version of the claim in Premise Three work is an argument showing that the parties to the Covenant would want to dispense with offensive rights while retaining defensive rights. Let us call a right offensive if it is an entitlement to perform an offensive act. And let us similarly call a right defensive if it is an entitlement to perform a defensive act. We can define offensive and defensive acts as follows. An action is offensive if it is the first of a causal series that infringes some right of another agent. And an action is defensive if it is an attempt to prevent an imminent infringement of right on the part of another agent of an offensive sort. So what Hobbes must show is that parties to the Covenant would think it advantageous to abandon the right to initiate infringements of right against other agents, but that they would wish to retain the right to prevent infringements of right against themselves.

It is clear from Hobbes' discussion in Chapter XIII why the parties would want to abandon offensive rights. The retention of offensive rights
would be particularly damaging to the effort to establish peace and security in civil society, for if anyone had a right to make use of me whenever his survival was incompatible with my continued survival, or had a right to make use of my possessions when his survival depended on doing so, I would live in continual fear and distrust, both for my bodily security and for the fruits of my labor. By abandoning the right to make offensive moves against one another, each person has implicitly gained the right to be free from bodily interference. The retention of defensive rights, then, makes sense as protection for the newly-created right to be free from the offensive attacks of others. Thus the reason for the retention of the right to self-defense, as I have construed it, might be that the parties to the original Covenant would regard it as in their interest to retain a self-help mechanism for protecting themselves against the offensive violations they have agreed to abandon. That is, retaining the right to self-defense is a way of protecting themselves against the increased vulnerability that the abandonment of offensive rights creates.

But there is something puzzling about the retention of the defensive right. For if everyone has abandoned a right of offensive attack, against whom would civil agents require a right of defensive response? An answer may lie in the fact that there is someone who retains the right to attack offensively, namely the sovereign. The sovereign has unlimited rights, since he has all the rights of natural persons and has given up none of them in civil society. All other individuals give up a portion of their infinite right in favor of him, and in this way the sovereign comes to have infinite right over them. Arguably, then, the right of self-defense is primarily necessary against the sovereign, since he is the only person or entity who retains an entitlement to use the kind of force against which a right of self-defense in civil society might be necessary.

This would explain why Hobbes focuses repeatedly on the relation of citizens to the sovereign in his discussion of self-defense. For example, he famously says:

If the sovereign command a man (though justly condemned) to kill, wound, or maim himself, or not to resist those that assault him, or to abstain from the use of food, air, medicine, or any other thing without which he cannot live, yet hath that man the liberty to disobey. 41

And this liberty to disobey is thought to hold even with respect to actions that the sovereign is perfectly entitled to take. The principle leads to quite strong results:

Upon this ground, a man that is commanded as a soldier to fight against the enemy, though his sovereign have right enough to punish his refusal with death, may nevertheless in many cases refuse without injustice, as when he substitueth a sufficient soldier in his place. . . . 42

© 2001 University of Southern California and Blackwell Publishers Ltd.
The right to self-defense thus has great importance for the relation of citizens to their sovereign. It makes it permissible, for example, not only to attack the sovereign when threatened by him, but also to shirk the obligations of war and other dangerous service, as long as one does not substantially impair the sovereign’s ability to govern.

If the primary purpose of the right to self-defense is protection against an all-powerful sovereign, how should we understand the exercise of the right against an ordinary citizen? For sometimes the right will be invoked against those who attack without offensive rights. I am inclined to think there is some redundancy in Hobbes’ account here. Private citizens who act offensively are violating the terms of their Covenant. Thus arguably the right to self-defense is not required as against them, since they have placed themselves in a posture of war towards the rest of civil society. They might be attacked as covenant-breakers rather than as initiators of rights violations. One would require a clearer sense of Hobbes’ view of covenant-breaking to know whether this would be adequate. It might turn out, for example, that the correct sanction for covenant-breaking is expulsion from civil society, rather than death. Further exploration of this problem, however, is beyond the scope of our discussion.

My suggestion, then, is that for Hobbes, the right of self-defense serves an important political function, whatever other functions it might serve. It emerges as the central protection citizens have against a government whose laws threaten, rather than advance, their interest in bodily security. Although there are virtually no restrictions on what the sovereign may legitimately do to a subject, if a person is immediately threatened by some act of the sovereign’s, the subject retains the right to resist. But one might then ask: why does what looks like a personal right of self-defense against the sovereign provide a political safeguard against abuse of power? Recall that Hobbes says the duty of subjects to the sovereign is extinguished when the sovereign is no longer able to protect them. Hobbes explicitly connects this right to disobey the sovereign with the inalienability of the right to self-defense:

The obligation of subjects to the sovereign is understood to last as long, and no longer, than the power lasteth by which he is able to protect them. For the right men have by nature to protect themselves, when none else can protect them, can by no covenant be relinquished.

In other words, it is the right of self-defense that grounds the entitlement to disobey the sovereign when he is unable to protect them. The subjects have a right to reject the sovereign’s authority if, for example, he were to refuse on a consistent basis to tend to their well-being. By retaining a right of self-defense, citizens give themselves the right to reject the sovereign’s authority in the event that he no longer rules with their welfare as
his primary guiding principle. In other words, by retaining the right to self-defense, they have retained a natural right of revolution.

Granted, it is not the sort of protection against tyranny that we moderns would normally insist on in our political theory. For Hobbes' right to self-defense is a mere liberty right, rather than a full-fledged claim right. That is, it is a right that places no one under a correlative duty of non-interference. Thus the sovereign is under no obligation to respect the right to self-defense that citizens have, at least on the face of it. One might then wonder why the retention of this right is of any significance to citizens at all, given that the extension of a non-correlative right of resistance would appear to be largely useless as a guarantee of political freedom.

I would argue there are at least three reasons the retention of the liberty right is significant. First, the existence of the right means that citizens do not themselves violate the duties they have in foro interno if they respond to the sovereign's legitimate use of force with force. That is, they do not violate their natural obligations by resisting the sovereign, and thus they do not sin against God in doing so. Second, citizens do not violate their contractual obligation to one another by rejecting the sovereign's authority under these circumstances, as they would were they to reject his authority in other cases. They do not violate this obligation, because the Covenant specifically excludes self-defense and the rights that follow from it. They are thus free to reject the sovereign's authority under the applicable conditions without suffering the expulsion from the ranks of civil society that a rejection of the sovereign's authority would normally entail.

Third, arguably the sovereign does have a duty to respect the right to self-defense of his citizens in foro interno, since it would be a violation of a law of nature to disrespect the right. This may seem a feeble duty from a political standpoint, since it is not one that limits the sovereign's infinite right over his subjects. But the sovereign's duty in this regard may be more robust than one might have thought. For the sovereign might violate the laws of nature were he to enact a law contravening the right of self-defense of his citizens, or alternatively, were he to enact a law allowing for the punishment of those who had acted in exercise of their natural right to self-defense. While Hobbes nowhere directly addresses the question, he does say that a sovereign who attempted to punish the innocent would violate several of the laws of nature. Among other things, he would violate the law of Equity, by which the sovereign along with its judges and magistrates are bound. Our concern is with a different case, since in the face of a law forbidding self-defensive actions, a person who killed another in self-defense would not be innocent. (By "innocent," Hobbes means someone who is not in fact responsible for a violation of civil law.) Nevertheless, Hobbes does appear to treat the passage and enforcement of laws abridging the natural rights of citizens as itself a
violation of the laws of nature. So while he apparently allows that positive laws can violate laws of nature and still be laws, Hobbes suggests that the sovereign would violate his duty to God were he to pass such a law. The general thought is well expressed in a passage from the Latin edition of *Leviathan*: “[H]e who has the Supreme power . . . can do no injury to his citizens, even though, by iniquity, he can be injurious to God.”46

Particularly in view of this third point, then, the right of self-defense plays an indirect role in ensuring the civil liberties of citizens. Because the right follows from the fundamental law of nature, the sovereign is under what we would think of as a moral obligation to respect it, even if there is no political obligation to citizens to respect the right as such. When we combine this thought with the right of revolution Hobbes so passingly suggests, it is not implausible to suppose that a sovereign who failed to respect the right of self-defense of his citizens would be a sovereign to whom citizens no longer owed their allegiance.

University of Pennsylvania Law School

NOTES

1 Professor of Law, University of Pennsylvania Law School. I am indebted to Geoff Brennan, David Copp, David Gauthier, Ross Harrison, Leo Katz, Pasquale Pasquino, and Michael Ridge for comments on earlier drafts of this article, as well as to an anonymous reviewer for *Pacific Philosophical Quarterly*. I am also indebted to the members of the Analytic Legal Philosopher’s Conference, the Thursday Seminar Series at the Australian National University and the Political Philosophy Colloquium at Princeton University.

2 Judith Jarvis Thomson, “Self-Defense,” *Philosophy and Public Affairs* 20, no. 4 (Fall 1991): 283–310. In Thomson’s case, the relevant right is the right not to be attacked by the person attacking, rather than the right to life.


4 Ibid.


6 Ibid.

7 This also explains the limitation of my right to self-defense to cases “that exclude him not from appealing for protection to the law established by it.”


9 Hobbes in fact speaks of an abandonment or a transfer of right, which elsewhere he makes a point of distinguishing. *Leviathan*, Ch. XIV, ¶ 7. Hobbes defines a transfer of right as “when [the transferor] intendeth the benefit thereof to some certain person, or persons.” Ibid. Transferring a right of self-defense to another would mean abandoning the right with respect to that person. Abandoning the right to self-defense would mean giving up the right with respect to all others. I shall use the two interchangeably, however, since the difference is not important for our purposes.

10 There is obviously a step—a transfer of right involves some good to the agent who transfers it—that belongs after the second premise. I treat it as an intermediate conclusion,
rather than constructing the argument with it as premise, because it appears to be composed of two elements more basic than it.

11 Gregory Kavka suggests that the argument is invalid. Gregory S. Kavka, *Hobbesian Moral and Political Theory* (Princeton: Princeton University Press, 1986), § 8.2. But the invalidity strikes me as Kavka’s own making, based on his unnecessarily redundant formulation of Hobbes’ argument. As I have shown, there is a perfectly valid way of constructing the argument, one that is closer to Hobbes’ text than the reconstruction Kavka offers. I wish, then, to consider its soundness, something Kavka thinks it unnecessary to do.

12 *Leviathan*, Ch. VI, ¶ 53.

13 Ibid., ¶ 54.

14 I defend this interpretation of benefit below.

15 Granted, if this was the thought behind the premise, Hobbes might have put the point more clearly by requiring that the transfer be *intentional*, rather than voluntary. I think, however, that this is roughly what he meant. When Hobbes speaks of voluntariness, he sometimes seems to mean that something must be an act, as opposed to a bodily movement, and sometimes that it must be intended by the agent who performs it. At any rate, the distinction is not important in this context.

A further point about this premise is worth noting. One might think that a contractarian theory that allows for sucit, rather than express, agreement is a theory that rejects this first premise. But many such theories still require a voluntary act to which the tacit consent is tied.


17 The formulation of this premise may still seem overly egoistic. Why, for example, should we limit rational agency to pursuit of ends that specifically benefit the agent? Under (ii), that is, we still seem to rule out purely altruistic behavior. But I think this formulation is sufficiently broad to be consistent with a moderate maximizing conception of agency, without restricting the maximizing agent to egoistic ends. Since I believe this is a fairly faithful account of Hobbes’ view of rational agency anyway, (ii) seems the preferable formulation of the second premise.

18 Hobbes also offers a more practical argument for the irrationality of conveying away the right to defend against these harms—namely that it is impossible to tell whether the person coming to wound you will not also kill you.

19 There might even be such a sum that would induce me to abandon the right to defend my life itself, since I might wish to gamble that I would not need to assert it if offered a sufficiently large amount of money.

20 *Leviathan*, Ch. XIV, ¶ 27.

21 Ibid.

22 *De Cive*, p. 129 (emphasis added).


24 *Leviathan*, Ch. XIV, ¶ 29.

25 That Hobbes has something of the sort in mind is borne out by the fact that he never considers the possibility of insincere promises on this point.

26 He compares failing to stick to one’s obligations to the notion of absurdity in logic: “For as it is there called an absurdity to contradict what one maintained in the beginning, so in the world it is called in justice and injury voluntarily to undo that which from the beginning he had voluntarily done.” *Leviathan* Ch. XIV, ¶ 7.

A PUZZLE ABOUT HOBBES ON SELF-DEFENSE

29 The term was first used by Ned McClennen. Edward F. McClennen, Rationality and Dynamic Choice (Cambridge: Cambridge University Press, 1990).
30 Leviathan, Ch. XV, ¶ 4.
32 Hampton, p. 65.
33 Leviathan, Ch. XV, ¶ 5.
34 I deliberately do not say that we can presume that others are rational, but only that we cannot presume that they are irrational. I make use of the more qualified claim because Hobbes seems in many places to allow that human beings will have streaks of irrationality, such as when they are motivated by vanity and the desire for glory. Moreover, as I explain below, the distrust Hobbes has for certain kinds of contracts is a state of nature seems to depend on his view that human beings cannot be presumed to be perfectly rational.
35 Gauthier, p. 136.
36 Not only must there be a kind of contract that is unavailable in a state of nature, but it must also be the case that the latter are important for reaping the benefits civil society brings.
37 Leviathan, Ch. XIV, ¶ 16.
38 Ibid., ¶ 17.
39 Hampton thus concludes that Hobbes in effect tells the fool that it is always rational to keep covenants, thus contradicting several chapters of argument to the effect that it is not rational to keep covenants in a state of nature. Hampton, p. 65.
40 Admittedly the idea of a first act in a causal series is problematic, since causal series have no natural beginnings and endpoints. But we might think of voluntary human actions as initiating causal series, at least in many cases, and thus a series of events will have a causal initiation at the point at which an agent acts and thus sets in motion a series of ensuing events. I shall not, however, attempt to make this more precise.
41 Leviathan, Ch. XXI, ¶ 12.
42 Ibid., ¶ 16
43 One might wonder in this connection whether citizens have the right to revoke the initial authorization of right to the sovereign, and also about when the sovereign exceeds his legitimate authority relative to his duty to follow the laws of nature. These interesting problems, however, are beyond the scope of our present inquiry.
44 Leviathan, Ch. XXI, ¶ 21.
45 Hobbes says that punishment of the innocent would violate three laws of nature: first, the law requiring men always to look for some future good, second, the law against ingratitude, and third, the law of equity. Ibid., Ch. XXVIII, ¶ 22.
46 Latin edition translation from Ch. XXI, ¶ 7.