Hobbesian Contractarianism, Orthodox and Revisionist

Gerald Gaus

In seeking to understand Hobbes's contractarianism we are faced not simply with the problem of interpreting Hobbes, but of characterizing social contract theory, and the divisions within it. The latter is itself a complex subject, about which commentators disagree. And once we accomplish that, we face debates about what type of contract theory Hobbes offers, and, indeed, whether Hobbes is best interpreted as a contract theorist at all. In this chapter I put aside most of these important and difficult questions and focus on the contrast between two recent interpretations of Hobbes's social contract theory, the orthodox and the revisionist. The orthodox interpretation is, as one would expect, familiar; it is also elegant and in many ways powerful. Yet I believe that the revisionist interpretation is a decisive advance in our understanding of the problems with which Hobbes grappled, and how his social contract seeks to cope with them.

1. The Orthodox Contract: Hobbes, the Theorist of Strategic Interaction

Both the orthodox and revisionist readings are best understood as families of interpretations, and as with all families, there are lively disputes within them. The orthodox interpretation was spurred by the David Gauthier's *The Logic of Leviathan* and reached its height in Gregory Kavka's *Hobbesian Moral and Political Theory* and Jean Hampton's *Hobbes and the Social Contract Tradition*.³ It is the accepted view among leading philosophers who build upon Hobbes for their own social contractactarian (or

contractarian-inspired) contributions to contemporary moral and political philosophy.⁴ It has four elements.

1.1 Hobbesianism as an Account of Trials and Tribulations of Instrumentalist Reasoners

It is commonplace to describe Hobbes as claiming that people are inherently "selfish,"

"egoistic," or are "rational egoists," though at least since the work of Bernard Gert it has

been generally recognized that Hobbes's texts do warrant the claim that he embraces

"psychological egoism" — i.e., the thesis that people are moved *only* by considerations of

self-interest. Kavka famously argues that Hobbesian theory is based on "predominant

egoism:" most of the people most of the time are much more strongly inclined to act on

self-interest rather than altruistic motives, though with regard to families, close associates

and friends, people may act altruistically.

However, even this seems more restricted than the orthodox interpretation requires: the core claim of this reading is that, for Hobbes, practical reason is simply instrumental. "Reason [on Hobbes's view] dictates action just insofar as an agent already has ends, ends that do not derive from reason....It is the transfer of motive force from end to means by right reasoning that reason's dictates consist." Hampton developed the most sophisticated interpretation of Hobbes along these lines, emphasizing that he does not advocate a simple, bare-bones, conception of instrumental rationality according to which, whatever one happens to desire, one acts rationally just because one effectively satisfies that desire. As she reads Hobbes, some passages are quite clear that people are pursing mistaken ends, such as those "not contented with their present condition, [who] are inclined to continue the causes of war and to stir up trouble and sedition." So it cannot

be the case that a person is rational simply because she effectively seeks her ends, whatever they may be. Instead, Hampton maintains that Hobbes embraces a version of what has been called "deliberative rationality." What a person has reason to do what is what will achieve not what he *actually wants*, but what he *would want*, at the time of acting, if he had undergone a process of "ideal deliberation" in which a) he knew all the relevant facts...b) his reasoning was free from distorting influences, and c) he was not affected by desires produced in him by pathological processes." Gauthier has insisted that a conception of deliberative rationality is fundamental to the entire contractarian project; the contractarian's aim is to show how this prosaic conception of reasoning leads each of us to endorse principles of social cooperation that benefit all, and which each is moved to act on in the light of her own use of his deliberative reason. 11

1.2 Deliberative Rationality and Game Theoretical Analysis

We need such principles because agents who are fully deliberatively rational find themselves in strategic interactions in which each is worse off than she might (in some sense) be.

The key idea is that in many situations, if each person chooses what, given the choices of others, would maximize her expected utility, then the outcome will be mutually disadvantageous in comparison with some alternative — everyone could do better. Equilibrium, which obtains when each person's action is a best response to the other's action, is incompatible with (Pareto-)optimality, which obtains when no one could do better without someone doing worse. Given the ubiquity of such situations, each person can see the benefit, to herself, of participating with her fellows in

practices requiring each to refrain from the direct pursuit of her own utility, when such mutual restraint is mutually advantageous....We may represent such a practice as capable of gaining unanimous agreement among rational persons who were choosing the terms on which they would interact with each other. And this agreement [says Gauthier] is the basis of morality.¹²

Thus it is said by many that the heart of Hobbes's analysis of the state of nature is that individuals are caught in a Prisoners' Dilemma, as in Figure 1.¹³

 Betty

 Attacks
 Doesn't Attack

 Alf
 3
 1

 Doesn't Attack
 1
 2

 Attack
 4
 2

FIGURE 1: HOBBES' STATE OF NATURE AS A PRISONERS' DILEMMA

Hobbes, of course, holds that the state of nature would be a war of each against all others. For Hobbes the basic problem is that humans are fundamentally equal in the crucial sense that anyone can kill anyone else. This equality gives rise to war, because each person is as hopeful as the next of obtaining her goals. But that means if they both want something, neither is apt to give way, since neither considers herself inferior to the other. ¹⁴ On this reading, then, because deliberative rationality (or, on the more common interpretation, predominant selfishness) requires each to maximize the achievement of his or her own ends, Alf and Betty both order the possible outcomes of interactions as: (1) I win,

conquer the other and take his/her possessions and secure myself against future attack, (2) we are at peace, (3) we are at war, (4) the other wins, conquers me, and takes my possessions. Now the key to the Prisoners' Dilemma is that fully deliberatively rational individuals will both end up with their third ranked option. In the language of game theory, "attack" is the dominant strategy of each: no matter what the other does, one does best by attacking. Suppose that Alf will attack; then Betty had better attack too, as she will get her third rather than last ranked outcome. Suppose Alf will not attack: then Betty will attack, since she can get her first rather than second ranked outcome. Both reason this way, both attack, and they end up at the war. On this analysis, so far from being the result of any defect of rationality, it is the very commitment of each to best pursing his or her goals that leads to war in Hobbes's state of nature.

It is hard to overstate the influence of this line of thought: the idea that the chief barrier to social cooperation are interactions with Prisoners' Dilemma characteristics remains at the heart of much moral, social, and political theory. Now what I have called "orthodox" Hobbesians overwhelmingly see the social contract as a way to overcome the Prisoners' Dilemma: by agreeing to obey the sovereign and giving him the power to punish those who attack, we change the game so that we achieve the peace/peace outcome, which is everyone's second choice. Thus it makes perfect sense that Hobbes's social contract (in commonwealth by institution) is not a contract between the subjects and the sovereign, but among the subjects, for the crucial thing is that we covenant with each other (as players in Prisoners' Dilemmas) to be bound by the dictates of the sovereign (which determine justice) and so avoid Prisoners' Dilemmas. The social

contract raises everyone from her third to her second choice, thus advancing everyone's interests over the state of nature.

In the orthodox tradition, intra-familial disputes often focus on whether an alternative game theoretical analysis is appropriate.¹⁷ Hobbes tells us that "in the nature of man, we find three principal causes of quarrel: first, competition; secondly, diffidence; thirdly, glory."¹⁸ The Prisoners' Dilemma reasonably models conflict through competition (after all, in a Prisoners' Dilemma one seeks to come out on top in a competition). On the other hand, conflict through fear seems better modeled by the Assurance Game of Figure 2.

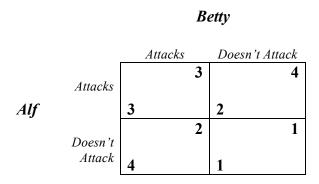


FIGURE 2: HOBBES'S STATE OF NATURE AS AN ASSURANCE GAME

Here Alf and Betty would most prefer to cooperate (not attack), but the worst thing is to be "suckered" — to cooperate when the other does not. Because there is a lack of trust in the state of nature, we fear each other. Not being confident about each other's motivations, we tend to take the less risky alternative of "attack" and so once again we end up in the attack/attack equilibrium. Although in this game peace (don't attack/don't attack) is the preferred equilibrium, Peter Vanderschraaf has shown that war easily comes about in a wide range of circumstances.¹⁹

1.3 The Social Contract as a Device for Pareto Superior Moves

The analysis of the strategic basis of war in the state of nature leads to the interpretation of Hobbes's contract as a device for mutual advantage. In both the Prisoners' Dilemma and the Assurance Game war is an equilibrium outcome that is worse for everyone than peace. Once we understand this, the point of the contract is clear: to provide for peace that renders each and every person better off. The use of each person's deliberative rationality leads her to see that she benefits from peace, and so she has reason to sign on to the creation of a sovereign who will secure a peaceful equilibrium.

The analysis of the main causes of war — whether a Prisoners' Dilemma, an Assurance Game, or some combination of the two²⁰ — leads to different understandings of the social contract. To the extent the analysis is based on endemic competition leading to Prisoners' Dilemmas, the aim of establishing a sovereign (by institution) is to create an actor with sufficient force to threaten effective punishment for attacking (or, more generally, defecting on the social contract; see §1.4). An effective threat changes the payoffs of the game such that attack-when-the-other-does-not-attack is no longer one's best outcome. The force of the sovereign can change the Prisoners' Dilemma interactions of the state of nature into an Assurance Game in society, in which because of the force of the sovereign we are confident enough to play "don't attack" (or, more generally, "keep the contract") option, leading to gains for everyone. On the other hand if, as does Hampton, one stresses the importance of the Assurance Game as the source of conflict in the state of nature, the Hobbesian contract is better understood as a coordination device in which each gains public assurance that the others will play their cooperative move, giving everyone their best outcome. Here cheating on the social contract is not a

Assurance Game: that was always the structure of the problem and the root of conflict. If so, what we all need is simply to be assured of each other's good intentions. To the extent one reads the core problem of Hobbes's social contract as *coordinating* on a sovereign to solve the Assurance Game in the state of nature it is not clear that Hobbes really offers a social *contract* — at least if we think of a contract as something that stops you from performing an action that in some sense you prefer to perform.²¹ If Hampton is correct, we all rank as best coordinating on a sovereign to provide order; if so, no one has incentive to defect from the contract since we get our first choice.

1.4 The Fundamental Importance of the Reply to the Fool

These first three foci of the orthodox view of the Hobbesian contract leads to the fourth: the overwhelming importance in Hobbes's theory of the reply to the fool. Recall:

The fool hath said in his heart, "there is no such thing as justice;" and sometimes also with his tongue, seriously alleging that "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." He does not therein deny that there be covenants, and that they are sometimes broken, sometimes kept, and that such breach of them may be called injustice, and the observance of them justice; but he questioneth whether injustice, taking away the fear of God (for the same fool hath said in his heart there is no God), may not sometimes stand with that reason which dictateth to every man his own

good; and particularly then, when it conduceth to such a benefit as shall put a man in a condition to neglect, not only the dispraise and revilings, but also the power of other men.

"The kingdom of God is gotten by violence; but what if it could be gotten by unjust violence? were it against reason so to get it, when it is impossible to receive hurt by it? and if it be not against reason, it is not against justice; or else justice is not to be approved for good."²²

The fool's challenge is fundamental if (*i*) according to Hobbes reason is essentially instrumental, deliberative, reasoning about how to best secure one's aims, and (*ii*) the social contract is simply a way to create an institution that allows a rational individual to best achieve her ends. We have seen that in Prisoners' Dilemmas and Assurance Games, each individual is better off if she somehow is locked into a cooperative (Don't Attack) strategy. This is especially striking in the Prisoners' Dilemma version, in which everyone's maximizing response makes everyone worse off. This leads to the enlightening and exciting idea of Gauthier's Hobbesian contractarian project: the whole edifice of social moral rules can be understood as way to secure Pareto-superior outcomes (those that make everyone better off than the state of nature). Mutual advantage is the foundation of morality, justice and, indeed, of political life.²³

The fool agrees with all of this but presses: if he knows that in some case the best response to others' observance of justice is to (in game theoretic terms) "defect" on the cooperative arrangement — aka act against justice — then since the *telos* of reason is to lead him to his good, and here his good endorses injustice, injustice is what reason

requires. The fool confronts the orthodox Hobbesian with a higher-level Prisoners' Dilemma, as depicted in Figure 3.

All Others

		aci we	ot enough t justly, so are in the state of nature	ac the	ough others et justly, so state of war s avoided
The Fool	Successfully acts unjustly	3	3	1	4
The Tool	Acts justly		1	-	2
		4		2	

FIGURE 3: ACTING JUSTLY AS A MULTI-PARTY PRISONERS' DILEMMA

Here we see the fool as playing against all others, and for him the relevant concern is whether enough others act according to the rules of justice so that the state of war in avoided. Again, because this is a Prisoners' Dilemma the fool's rational strategy is to defect; recognizing this, it seems that we all should be fools.

The fool is thus the core challenge on the orthodox view of the Hobbesian contract, and the crux of orthodox interpretations of Hobbes, as well as orthodox contemporary Hobbesian views such as Gauthier's, is to reply to the fool. We can identify at least three important types of replies.

(*i*) Following Hobbes, we may say that the fool is never rational to assume that he can achieve his favored payoff (others act justly while the fool successfully acts unjustly); if that is not a rational option, injustice cannot be rational dominant strategy. Hobbes seems

to accept that the fool's reasoning may be acceptable in the state of nature, but in civil society it is "fallacious."

He....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 therefore, if he be left, or cast out of society, he perisheth; and if he live in society, it is by the errors of other men, which he could not foresee nor reckon upon; and consequently [he has acted] against the reason of his preservation, and so, as all men that contribute not to his destruction forbear him only out of ignorance of what is good for themselves.²⁴

Hobbes seems to argue here that the fool's action cannot be rational even if he succeeds, because the belief upon which his action depends — that he can continue to deceive others — is not well-grounded. And so the fool is playing an irrational game with his life. Note, though, that this seems to assume that the fool is an unusual reasoner: if we all reason like the fool, then we cannot expect a society to maintain itself and so there really is no possibility of stable social life.

(ii) This suggests a second reply: because we all see that if we reasoned like the fool we could not obtain the benefits of social life, and because these are great benefits that we all require, we can see that what is "fallacious" in the fool's reasoning is the manner in which he instrumentally reasons. The fool seems to engage in what game theorists call "subgame perfect" reasoning: at each point in his decision tree he asks

himself "given where I now am, what action will best promote my ends?" Gauthier has argued that this is a sort of myopic instrumentalism will prevent the fool from achieving his ends. He employs an example from Hume:

My crops will be ready for harvesting next week, yours a fortnight hence. Each of us will do better if we harvest together than if we harvest alone. You will help me next week if you expect that in return I shall help you in a fortnight. Suppose you do help me. Consider my decision about helping you. I have gained what I wanted — your assistance. Absent other not directly relevant factors, helping you is a pure cost to me. To be sure, if I were to help you I should still be better off than had I harvested alone and not helped you, but I should be better off still if having received your help, I did not return it. This calculation may appear short sighted. What about next year? What about my reputation? If I do not help you, then surely I shall harvest alone in future years, and I shall be shunned by our neighbors. But as it happens I am selling my farm when the harvest is in and retiring to Florida, where I am unlikely to cross paths with anyone from our community.

Being rational persons, we both know this, the scenario I have sketched is one each of us can sketch — and each of us knows it to be true. It would be pointless of me to pretend otherwise. So you know that I would not return your help, and being no sucker, will therefore leave me to harvest my crops alone. Neither of us will assist the other, and so each of us will do worse than need be. We shall fail to gain the potential benefits of cooperation.²⁵

The problem can be depicted as in Figure 4. My neighbor chooses at the diamond, I choose at the ovals; payoffs are ordered from 1 (best) to 4 (worst), first my neighbor's, then mine. The problem is that my neighbor knows the decision tree, and that I am "subgame perfect" rational (I will always take the branch of the decision tree that gives me the highest payoff). Once it is my turn to choose I will look to what decision will be best for me from here on into the future. If my neighbor helped, I do best by not helping (getting 3 rather than 4). If my neighbor didn't help, I do best by not helping (getting 1 rather than 2). As in the Prisoners' Dilemma my dominant strategy is not to help. My neighbor knows this, and so will not help; once again we are both stuck at a Pareto-inferior outcome.

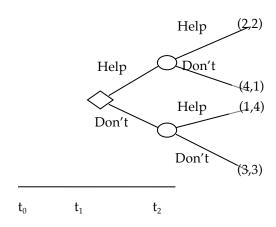


FIGURE 4: HUME-GAUTHIER ONE-PLAY HARVESTING GAME

Being a "subgame perfect" reasoner, once the fool gets to t_2 , and has received help, his best strategy is not to help in turn. But if Hobbes is right, the fool cannot reliably fool the neighbor, so the neighbor will not help, and so no one will get her crops in. Gauthier is convinced that an adequate conception of instrumental rationality cannot lead to this

result. The aim of instrumental rationality is that one's life goes "as well as possible," but here it instructs one to do what does not make one's life go well. A broader, better, conception of instrumental rationality would require the action that is part of a plan that makes one's life go as well as it can, and that plan would require following through with one's commitment in this case. So the fool's reasoning is fallacious not just in assuming that he can reliably fool others, but in thinking that the way of reasoning based on his own good requires decision-by-decision optimization rather than whole-life optimization.

(*iii*) A great attraction of Hampton's analysis of Hobbes's contract is that, rather than depicting it as a multi-person Prisoners' Dilemma (in which the fool seems to identify the best strategy), it is understood as a type of coordination game, in which each party ranks as best escaping the state of nature by coordination on a sovereign.²⁷ Or, more strictly, Hampton maintains that Hobbes's social contract by institution can be analyzed as a series of games (or stages), which includes a game of coordinating on a particular sovereign and a further game of empowering the sovereign with sufficient force to secure peace. In an importance sense, the fool misunderstands the game.

2. Revisionist Interpretations: Hobbes, the Theorist of Public Reason

Although we have seen that proponents of the orthodox interpretation of Hobbesian contractarianism disagree about much, as an overall approach it contains shared dominant themes. What I shall call "revisionist interpretations" are more diverse — they do not share the awareness of a common approach that we find in Gauthier, Hampton, and Kavka. Nevertheless, we can discern ssome common themes that connect the work of the

two main revisionists, S.A. Lloyd and R.E. Ewin, as well as others who take this route. The nature and coherence of the revisionist view is nicely brought out by contrasting their analysis to the orthodox tradition on the four crucial elements we have considered: the nature of rationality, the dynamics of conflict in the state of nature, the point of the contract, and the analysis of Hobbes's reply to the fool.

2.1 A Richer Conception of Rationality

A revisionist point of departure is Hobbes's repeated claim that the laws of nature can be summed up by the Golden Rule²⁸ — "Do not that to others, you would have done to thyself:"²⁹ Kavka famously holds that Hobbes really intended to present a "Copper Rule:" "Do unto others as they do unto you."³⁰ This is very much like the tit-for-tat rule advocated by Robert Axelrod as a strategy in iterated Prisoners' Dilemmas, and makes perfect sense on the orthodox analysis.³¹ But to embrace the Copper Rule interpretation is to hold that Hobbes was confused about the very point of his laws of nature. The tenth law is entirely explicit in regard to peace: "If in ... the making of peace, men require for themselves that which they would not have to be granted to others, they do contrary to the precedent law, that commandeth the acknowledgement of natural equality, and therefore also against the law of nature."³²

Lloyd has trenchantly argued that, in addition to instrumentalism or deliberative rationality, Hobbes's understanding of the demands of reason require justifications that meet a condition of *reciprocity*.

From our common definition of man as rational, Hobbes argues that we won't count a man as rational unless he can formulate and is willing to offer, at least *post hoc*, what

he regards as justifying reasons for his conduct (and beliefs). But to offer some consideration as justifying one's action commits one to accepting that same consideration as justifying the actions of others, *ceteris paribus*....So one acts against reason when one does what one would judge another unjustified in doing.³³

Lloyd advances what she describes as Hobbes's "reciprocity theorem:"

- 1. Man is rational.
- 2. Insofar as man is rational, his action is not contrary to reason.
- 3. That which is not contrary to reason is judged to be done with right. But because what is judged to be done without right is not judged to be done with right, it follows (by contraposition) that whatever one judges to be done without right is contrary to reason; and so that
- 4. To do what one judges to be done without right is to act contrary to reason.
- 5. If one judges another's doing of an action to be without right, one judges the action done to be done without right.
- 6. Therefore, If one judges another's doing of an action to be without right, and yet does that action oneself, one acts contrary to reason (from 4 & 5). That is, to do what one condemns in another is contrary to reason.³⁴

On this analysis one cannot with reason do what one objects to another doing: if Alf seeks peace and objects to Betty's unilateral defection in a Prisoners' Dilemma as preventing them from leaving the state of nature, then his own unilateral defection is against reason. It is, *pace* the instrumentalist account, not the rational thing to do.³⁵ Ewin, though he employs the language of virtue, also holds that Hobbes sees a reasonable

social individual as one who does not insist on always getting her maximal outcome. A reasonable Hobbesian only demands a "fair go," and does not seek to always come out on top.³⁶

We may, indeed, question whether instrumentalism is really the heart of Hobbes's conception of practical rationality. In Chapter V (on "Reason in General") of *Leviathan*, Hobbes characterizes theoretical reason in terms of a reckoning of the "consequences of general names;" irrationality involves "absurdity" – senseless speech arising out of a contradiction in the use of names. In Chapter XIV, Hobbes draws on this idea in explaining why one should keep one's covenants: an injustice arising out of a broken contract

is somewhat like to that in which in the disputations of scholars is called absurdity. For as it is there called an absurdity to contradict what one maintained in the beginning, so in the world it is called injustice and injury voluntarily to undo that from the beginning he had voluntarily done.³⁷

Hobbes thus suggests a conception of rational action as a consistency relation between an utterance at time t_1 about one's future will to perform act ϕ at time t_2 and ϕ -ing at time t_2 . So Consider again Gauthier's harvesting game (§1.4). If at time t_0 I will to help my neighbor at t_2 in return for his help at t_1 , then I should help at t_2 . If I do not help, it is as if my action at t_2 contradicted my will at t_0 and so I committed a sort of absurdity — declaring at t_0 my will to help at t_2 and then by my action at t_2 declaring that was not my will. If I did not understand my words at t_0 to indicate the state of my will at t_2 , I "should

not have let them run."39 Notice that Hobbes thus links this sort of absurdity in action to absurdity in speech, and so back to his core doctrine of reason in Chapter V of Leviathan. It is crucial to realize that justice derives from contracts, 40 and contracts are about words; the performance of a contract is an act consistent with one's earlier words about the content of one's will. To be sure, if some intervening event greatly alters the situation, then my will can rationally change between t₀ and t₂, but if no new information arises between t₀ and t₂ such that, if I had the information at t₀ I would not have declared my will to help at t2, then I am rationally bound by my declaration at t0. "For that which would not hinder a man from promising, ought not to be admitted as a hindrance to performing."41 By appealing to the criterion of utterance-action consistency, Hobbes upholds the rationality of performing one's part of a covenant in the state of nature when the other has already has done his part. 42 If Hobbes's conception of a person who acts in accord with reason cannot be reduced to instrumentalism or deliberative rationality, the state of nature cannot be analyzed in terms of the trials and tribulations of maximizers and their strategic dilemmas.

2.2 Morality and Private Judgment in the State of Nature

On the orthodox account, the state of nature is a condition of unconstrained maximization not simply because reason is purely maximizing, but because orthodox maximizing agents act in what is essentially a moral vacuum. To be sure, orthodox views recognize the laws of nature, but they are almost always understood as mere maxims of prudence: good guides as to how a person can best secure her life or, perhaps more generally, her basic interests. "The laws of nature," Kavka tells us, "are grounded in self-interest and

may be considered rules of rational prudence." On the orthodox view all real restraint derives from the social contract.

Lloyd compelling argues against this orthodox interpretation, maintaining that the laws of nature are not plausibly understood as mere maxims that point to the good of each and every person, but are rules that instruct us how to secure the common good of a peaceful and cooperative social life. The problem with the "maxims of self-interest" interpretation favored by the orthodox view is that, while it is clear that the laws of nature are requirements of peaceful social life, Hobbes recognizes that peace is not in everyone's interest: "needy men, and hardy, not contented with their present condition; as also, all men that are ambitious of military command, are inclined to continue the causes of war; and to stir up trouble and sedition: for there is no honour military but by war; nor any such hope to mend an ill game, as by causing a new shuffle."44 To make sense of such cases the orthodox view must insist that such men are deliberatively irrational, and do not really endorse such ends that conflict with peace; only ends that are consistent with self-preservation are admissible. 45 But, if this is so, then it is hard to see how Hobbes can advocate the twentieth law of nature, "to protect in war the authority, by which he is himself protected in time of peace."46 The point of all the laws of nature are requirements for a peaceful and prosperous social life that conform to the reciprocity theorem. Hobbes's moral theory identifies the conditions for "peaceful life in community."47

We might wonder, then, why the state of nature would be a state of war. If reason requires reciprocity, and the laws of nature sum up the requirements of reciprocity concerning actions conducive to a peaceful and prosperous social life, from whence

comes the conflict? To agree on the rules for peaceful social life does not, alas, itself produce such a life, even among rational individuals who honor reciprocity. For one, the Assurance Game (§1.2) still is relevant. The laws of nature only provide conditional guidance;⁴⁸ if you do not believe that others will respect you, you are justified in acting on your right of nature, "the liberty each man hath, to use his own power, as he will himself, for the preservation of his own nature, that is to say, of his own life; and consequently, of doing any thing, which in his own judgment, and reason, he shall conceive to be the aptest means thereunto." Even if all are reasonable, it may not be common knowledge that all are reasonable, and so each may attack out of fear the other is not reasonable. And, of course, that reason instructs us to conform to the law of nature by no means ensures that we will do so, even when others do, for the "without the terror of some power to cause them [the laws of nature] to be observed, are contrary to our natural passions, that carry us to partiality, pride, revenge, and the like." ⁵⁰

Even if we all know that we all accept and will comply with the laws of nature, conflict would remain pervasive, for in the state of nature each is guided by her private judgment as to the demands of morality, as expressed by the laws of nature.⁵¹ In a crucial passage in Chapter V of *Leviathan*, in his initial discussion of reason, Hobbes writes:

reason itself is always right reason, as well as arithmetic is a certain and infallible art, but no one man's reason, nor the reason of any one number of men, makes the certainty... And therefore, as when there is a controversy in an account, the parties must by their own accord, set up, for right reason, the reason of some arbitrator, or judge, to whose sentence they will both stand, or their controversy must either come to blows, or be undecided, for want of a right reason constituted by nature, so is it

also in all debates of what kind soever. And when men that think themselves wiser than all others, clamour and demand right reason for judge, yet seek no more, but that things should be determined, by no other men's reason but their own, it is as intolerable in the society of men, as it is in play after trump is turned, to use for trump on every occasion, that suite whereof they have most in their hand. For they do nothing else, that will have every of their passions, as it comes to bear sway in them, to be taken for right reason, and that in their own controversies, bewraying their want of right reason, by the claim they lay to it. ⁵²

Because, "all laws, written, and unwritten, have need of interpretation," when, as in the state of nature, we each employ our private judgment, we disagree. "The unwritten law of nature, though it be easy to such, as without partiality and passion, make use of their natural reason, and therefore leaves the violators thereof without excuse; yet considering there be very few, perhaps none, that in some cases are not blinded by self-love, or some other passion; it is now become of all laws the most obscure, and has consequently the greatest need of able interpreters." Those who insist on employing their own reason to determine the law of nature, asserting that their reason is right reason, prevent a peaceful social life, for they are essentially insisting that they remain in the state of nature. Such individuals, as Hobbes says in the above quoted passage, are "intolerable in the society of men" for "they do nothing else, that will have every of their passions, as it comes to bear sway in them, to be taken for right reason."

2.3 The Social Contract as the Construction of Public Reason

On this interpretation, the crux of Hobbes's social contract is not a power that can induce maximizers to cooperate, or even simply one that assures all of the cooperative intentions of the rest: it requires an alienation of each of her right of private judgment and the erection of a commonwealth whose laws constitute public reason. In game theoretic terms, the social contract is essentially a coordination device, providing "a public mark of right reason" 55 so as to avoid conflict and allow for peaceful coordination and cooperation. Each alienates his own right to private judgment, on the condition that others do so, by settling on a sovereign, whose voice is the voice of public reason: "we are not every one, to make our own private reason, or conscience, but the public reason, that is, the reason of God's supreme lieutenant, judge."56 Hobbes thus proposes that disagreements in private reason (including disputes about the demands of the laws of nature) are to be resolved by the sovereign, who is to serve as an arbitrator. Ewin stresses this procedural aspect of Hobbes's proposal: even if we cannot agree on the merits of substantive doctrines or opinions, we may be able to resolve disputes through appeal to a judge or umpire, who provides a public decision about what to do. 57 The conflict between the conclusions of our practical reasoning can be resolved by following the practical reason of the judge.

Thus the social contract: "A *commonwealth* is said to be *instituted*, when a *multitude* of men do agree, and *covenant*, *every one*, *with every one*, that to whatsoever *man*, or *assembly of men*, shall be given by the major part, the *right* to *present* the person of them all (that is to say, to be their *representative*) every one, as well he that *voted for it*, as he that *voted against it*, shall *authorize* all the actions AND JUDGMENTS, of that man, or

assembly of men, in the same manner, as if they were his own, to the end, to live peaceably amongst themselves, and be protected against other men." The idea of accepting and authorizing *the judgment* of the sovereign occurs repeatedly in *Leviathan*. It is important that for Hobbes that while the *will* concerns deliberation about action, *judgment* is "the last opinion in the search of the truth." Hobbes identifies public reason with judgment of the sovereign, and so the sovereign provides a public determination of the truth, for example, of a claim that a miracle has occurred. Certainly Hobbes thinks that "when it comes to confession of that faith, *the private reason must submit to the public*." His position seems to be, though the sovereign cannot directly command us to believe a proposition since belief is not under the control of the will, 62 (i) the sovereign has authority to declare public truths from which we are obligated not to dissent; (ii) we can be obligated to publicly affirm these truths; and (iii) the sovereign has authority to shape the environment in which opinions are formed.

This "umpiring" or procedural solution to disagreement seems uncontroversial enough when applied to political disputes; however, given Hobbes's analysis it applies to all moral disputes as well. Remember, Hobbes starts out with disputes about the laws of nature — basic rules of ethical social conduct. His solution is to politicize all disputes about interpretations of these rules of social conduct by submitting them to the sovereign. If we closely follow Hobbes, it looks as if the political procedure will be determinative of all moral disputes about interpersonal conduct — having justified a judge or an umpire, we appear to have reason to appeal to it when we disagree about the dictates of the basic rules of ethical social conduct. Politics seems to swallow up morality. Thus Hobbes is scathing about the doctrine "That every private man is judge of good and evil actions."

While this "is true in the condition of mere nature, where there are no civil laws; and also under civil government, in such cases as are not determined by the law. But otherwise, it is manifest, that the measure of good and evil actions, is the civil law; and the judge the legislator, who is always representative of the commonwealth."⁶⁴ This is not to say that the sovereign's judgment determines the truth about the laws of nature and whether they have been violated: Hobbes is clear that the sovereign can err in interpreting the laws of nature. ⁶⁵ It is to say, though, that according to the contract we no longer have the right to judge him to be wrong, or to act contrary to his judgment. In contrast, the sovereign does construct the truths of justice; justice is determined by the civil law, and there is no civil law without the sovereign making it so. ⁶⁶

It is easy to see this as simply an alienation contract, in which each gives up the right to judge for herself, but it is better understood as also a coordination agreement, in which, each seeing the need for a shared interpretation of natural and civil law, converges on a common view — that of the sovereign. It does not matter so much who is the voice of public reason, so long as there is one.

Just as in the orthodox interpretation there are more absolutist and more "Lockean" readings of the contract, so too on the revisionist view. Lloyd defends the Lockean position. ⁶⁷ As she understands Hobbes, if we conceive of the state of nature as a condition under which people are guided by private judgment rather than public reason, the contrast between the state of nature and the civil condition becomes a continuum. On the one end is complete private judgment, on the other total submission to public reason. Now, she suggests, to avoid the first we need not rush into the second.

Because our motive, as rational agents, in transferring a portion of our right of private judgment to the public is to increase the effectiveness of our agency, we have no reason to transfer more to the public than is optimal for securing that end, and will indeed, if we are rational, resist pressures to do so. It is in this way that Hobbes sets the proper exercise of sovereign power, and the desirable range of subjects' right, much closer to what we would judge to be proper than to the authoritarian extreme generally attributed to Hobbes.⁶⁸

Ewin reads Hobbes as aiming at a contract that totally alienates the right of private judgment, but failing to achieve it: rights to private judgment "kept turning up" throughout his theory to the extent that, in the end, Ewin argues that Hobbes's subjects use their private reason to determine whether the sovereign is adequately performing his tasks.⁶⁹

It is hard not to conclude that such an incomplete alienation would be a disastrous outcome in Hobbes's eyes, for he saw the insistence on a right of private judgment as one of the great threats to civil order. Hobbes insists that *whatever* the sovereign proclaims is public reason *is* public reason and no limits can be placed on the sovereign's authority: "he is judge of what is necessary for peace; and judge of doctrines: he is sole legislator; and supreme judge of controversies." Hobbes adamantly denies, then, that the sovereign's authority can effectively be limited within some range of reasonable disputes; we cannot say that for all disputes within some range the sovereign is definitive, but outside of this range he has no authority. For Hobbes, identifying any such range will engender new disputes: some will employ their private reason to say that the sovereign has acted outside his authorization, and so set up his own private judgment in opposition

to the sovereign. This is precisely the problem Hobbes sought to avoid: each is insisting that "things should be determined by no other men's reason but their own." If you set yourself up in opposition to the sovereign, and insist that you shall be guided by your private reason, there will arise disputes, controversies, and at last war. Thus, any time that you dispute the sovereign's claim that he is acting within the range of legitimate disputes, this *ipso facto* is the type of dispute in which the clash of reason endangers society. Hobbes is convinced that such disputes are not only dangerous, but likely. Disagreement resulting from the use of private reason is endemic: there is no matter about which we can be confident that people will not disagree.

3.4 The Fool as a Fool, not just Mistaken

On the orthodox interpretation the fool's challenge is fundamental: reasoning is entirely instrumental, and the point of the social contract in establishing justice is to advance each agent's ends. So the fool raises the obvious question: if his instrumental reasoning shows that breaking a contract — including the social contract by rebelling — best promotes his ends, he can have no reason not to do so. The orthodox interpretation takes the fool so seriously because he is not really very foolish; he has the right conception of reason and the social contract, and so presses what seems to be a pretty obvious implication.

Certainly the orthodox interpretation is correct in holding that Hobbes believes that the fool makes an error in reckoning what best serves his interest; the fool, Hobbes tells us, acts imprudently *and* against reason.⁷¹ And here lies the rub: to act against prudence appears distinct from acting against reason. The orthodox interpretation, since it equates reason and prudence cannot make sense of this. But the revisionists see prudence as,

essentially, only an aspect of reason. The crux of the revisionist analysis of the fool is that Hobbes holds that the fool manifestly acts against reason holding that reason favors injustice, but Hobbes seeks to answer him in his own terms.⁷² Even on his impoverished conception of rationality he is mistaken. The fool is really a mistaken fool. He does not understand what reason is, and gets it wrong even on his erroneous conception. So the reply to the fool is by no means the centerpiece of *Leviathan*'s analysis.

An adequate analysis of the reply to the fool must give pride of place to the fact that the fool is not denying that the *laws of nature* generally bind him when they are not to his advantage: the fool is questioning the third law of nature, that "men perform their covenants made."⁷³ This law is the lynchpin of justice and, indeed, the entire social contract, for on Hobbes' view justice is conceptually tied to covenants. By attacking the third law of nature, the fool holds that reason endorses rebellion when rebellion is in his interests. So the fool's failure to grasp reason must derive from something special about the rationality of contracting, as opposed to the restraints of the laws of nature generally. The fool appears at the outset of chapter 15; at the end of chapter 14 Hobbes presented the utterance-action consistency model of rationality (§2.1), which claims that reason binds one's present action to prior states of the will (via speech), and that is why one of obligated to keep one's covenants. The fool fails to see this, and indeed adopts the pure maximizing theory of rationality. He thus cannot see any reason to keep a covenant when it is not in his interest, and so cannot see how reason endorses justice. The fool is willing to act in absurd ways in order to advance his interests. A fool indeed.

3. Conclusion

In different ways the orthodox and revisionist readings of Hobbes's social contract show how it is an elegant and in some respects compelling solution to the problem of social order in a world of individual conflict. And reading Hobbes as the orthodox view recommends is an aesthetic delight, presenting an elegant account of morality and politics based on a simple model of human rationality. But unless one is an economist or a decision theorist, the parties to the orthodox social contract seem alien, unreasonable, creatures, relentlessly focused on their own ends with little concern for treating others fairly. In contrast, the revisionist Hobbes speaks to our fundamental concern: how individuals who apply their reason and arrive at diverse and conflicting views of morality and politics can come to endorse a common, just, political order and a set of laws. "The central idea of' Hobbes, wrote Ewin, "is that of disagreement. People have different ideas of what is valuable or reasonable, and therefore different ideas about what natural law requires or allows." 74 The fundamental task of the social contract is thus an agreement among all us to coordinate on a specific voice of public reason. Hobbes's problem remains our problem, even if we recoil at his solution.

Notes

¹ It is not even clear when it begins. Many trace it back to the ancient world. See John Thrasher, "Reconciling Justice and Pleasure in Epicurean Contractarianism," *Ethical Theory and Moral Practice*, forthcoming. Cf. David Boucher and Paul Kelly, "The Social

Contract and its Critics: An Overview" in their edited collection, *The Social Contract from Hobbes to Rawls* (London: Routledge, 1994), pp. 1-34.

- ² The classic interpretation of Hobbes as a natural law account is Howard Warrander's *The Political Philosophy of Hobbes: His Theory of Obligation* (Oxford: Clarendon Press, 1957).
- ³ David Gauthier, *The Logic Of Leviathan* (Oxford: Clarendon Press, 1969); Gregory Kavka, *Hobbesian Moral and Political Theory* (Princeton: Princeton University, 1986); Jean Hampton, *Hobbes and the Social Contract Tradition* (Cambridge: Cambridge University, 1986). 1986 was the greatest vintage for the orthodox view.
- ⁴ For example, James Buchanan, *The Limits of Liberty: Between Anarchy and Leviathan* (Chicago: University of Chicago Press, 1975); Loren Lomasky, *Persons, Rights, and the Moral Community* (Oxford: Oxford University Press, 1987), chap. 4; David Gauthier, *Morals by Agreement* (Oxford: Oxford University Press, 1986). Jan Narveson, *The Libertarian Idea* (Philadelphia: Temple University Press, 1988), chap. 12.
- ⁵ For an extended treatment see Robert Shaver, *Rational Egoism: A Selective and Critical History* (Cambridge: Cambridge University Press, 1999), chap. 2.
- ⁶ Bernard Gert, "Introduction" to *Man and Citizen* (Garden City, NY: Anchor Books, 1972), esp. pp. 5-14. See also F. S. McNeilly, "Egoism in Hobbes," *Philosophical Quarterly*, vol. 6 (July 1966): 193-206.

⁷ Kavka, *Hobbesian Moral and Political Theory*, pp. 64-5.

- ⁸ Stephen Darwall, *The British Moralists and the Internal 'Ought'': 1640-1740* (Cambridge: Cambridge University Press, 1995), pp. 58-9.
- ⁹ Hobbes, *Leviathan*, edited by Edwin Curley (Indianapolis: Hackett, 1994), chap 11, ¶4 (p. 59).
- ¹⁰ Hampton, *Hobbes and the Social Contract Tradition*, p. 40. Emphasis omitted.
- ¹¹ David Gauthier, "Why Contractarianism?" in *Contractarianism and Rational Choice*, edited by Peter Vallentyne (Cambridge: Cambridge University Press, 1991), pp. 15-30.
- ¹² Ibid., pp. 22-3.
- ¹³ For an excellent analysis, which is far more sophisticated than that sketched here, see Kavka, *Hobbesian Moral and Political Theory*, pp. 109ff.
- ¹⁴ Hobbes, *Leviathan*, chap. 13, ¶¶1-8 (pp. 74-6).
- ¹⁵ In addition to Gauthier's work, see Cristiana Bicchieri, *The Grammar of Society: The Nature and Dynamics of Social Norms* (Cambridge: Cambridge University Press, 2006), chap. 1.
- ¹⁶ Hobbes, *Leviathan*, chap. 17, ¶13 (p. 109).
- ¹⁷ See Hampton, *Hobbes and the Social Contract Tradition*, chap. 2.
- ¹⁸ Hobbes, *Leviathan*, chap. 13, ¶6 (p. 76).
- ¹⁹ Peter Vanderschraaf, "War or Peace? A Dynamical Analysis," *Economics and Philosophy*, vol. 22 (2006): 243-79.

- ²⁰ Kavka proposes an "assurance dilemma" that combines both. Vanderschraaf provides a sophisticated analysis of this complex game in ibid.
- ²¹ See Hampton, *Hobbes and the Social Contract Tradition*, pp. 187-8; Gauthier disagrees in "Hobbes's Social Contract," *Nous*, vol. 22 (1988): 71-82. Hampton's idea of a self-enforcing contract (qua a coordination agreement) is to be distinguished from that proposed by Ken Binmore, in which the self-enforcing contract includes the possibility of punishment for violation. See his *Natural Justice* (Oxford: Oxford University Press, 2006), esp. chaps. 4 and 5.

²² Hobbes, *Leviathan*, chap 15, ¶4 (p. 90).

²³ For a survey and criticism of this idea, see Russell Hardin, *Indeterminacy and Society* (Princeton: Princeton University Press, 2003), chap. 4.

²⁴ Hobbes, *Leviathan*, chap 15, ¶5 (p. 92).

²⁵ David Gauthier, "Assure and Threaten," *Ethics*, vol. 104 (July 1994) p. 692. See David Hume, *Treatise of Human Nature*, 2nd ed., edited by L. A. Selby-Bigge and L. P. H. Nidditch (Oxford: Clarendon Press, 1976), Book III, Part ii, §5.

²⁶ Gauthier, "Assure and Threaten," p. 694.

²⁷ See Hampton, *Hobbes and the Social Contract Tradition*, chap. 2, esp. p. 65.

²⁸ S. A. Lloyd, *Morality in the Philosophy of Thomas Hobbes: Cases in the Law of Nature* (Cambridge: Cambridge University Press, 2009), p. 14.

²⁹ Hobbes, *Leviathan*, chap. 15, ¶35 (p. 99). Emphasis in original.

- ³⁰ Kavka, *Hobbesian Moral and Political Theory*, p. 347. He deems it the "Copper Rule" "because it glitters less brightly as an inspiring ideal of moral conduct than does the Golden Rule."
- ³¹ Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984), chap.
- 2. According to tit-for-tat, you should cooperate with another player on the first move and then "do unto him on the *nth* move what he did unto you on the *nth*–1 move." For an informed analysis, see Binmore, *Natural Justice*, pp. 77-9.
- ³² Hobbes, *Leviathan*, chap. 15, ¶22 (p. 97).
- ³³ Lloyd, *Morality in the Philosophy of Thomas Hobbes*, p. 4.
- ³⁴ Ibid., pp. 219-20. Emphasis in original, citations omitted.
- ³⁵ Game theorists abhor this line of argument; Binmore attributes it to Kant, and so holds that Kant's only real claim to "fame as a moral philosopher is based merely have invented one of the fallacies of the Prisoners' Dilemma before anyone else…." *Natural Justice*, p. viii.
- ³⁶ R. E. Ewin, *Virtue and Rights: the Moral Philosophy of Thomas Hobbes* (Boulder, CO: Westview, 1991), pp. 192, 202.
- ³⁷ Hobbes, *Leviathan*, chap. 14, ¶7 (p. 81)
- ³⁸ See further Philip Pettit, *Made with Words: Hobbes on Language, Mind, and Politics* (Princeton: Princeton University Press, 2008), pp. 60-66.
- ³⁹ Hobbes, *Leviathan*, chap. 15, ¶15 (p. 83).

- ⁴⁰ In the table of sciences in chapter 9 of *Leviathan*, justice is a consequence of speech, whereas ethics (the laws of nature) are a consequence of the passions. Thus injustice involves absurdity in a way inequity need not. This, we will see (§2.4) is important in the reply to the fool.
- ⁴¹ Hobbes, *Leviathan*, chap. 14, ¶20 (p. 85).
- ⁴² See ibid., chap. 14, ¶¶18–20 (pp. 84–5). See also ¶27 of chapter 14. Gauthier suggests an account very much like this; Gauthier's later work has affinities with important revisionist themes. See my *Order of Public Reason* (Cambridge: Cambridge University Press, 2011), pp. 79-86. See also Ewin, *Virtue and Rights*, pp. 113-6.
- ⁴³ Kavka, *Hobbesian Moral and Political Theory*, p. 309.
- ⁴⁴ Hobbes, *Leviathan*, chap. 11, ¶4 (p. 58); Lloyd, *Morality in the Philosophy of Thomas Hobbes*, p. 121.
- ⁴⁵ Recall that this is Hampton's view of such people (§1.1).
- ⁴⁶ Hobbes, *Leviathan*, Review and Conclusion, ¶5 (p. 490). Emphasis in original.
- ⁴⁷ Ewin, Virtues and Rights, p. 181.
- ⁴⁸ Hobbes, *Leviathan*, chap. 15, ¶36 (p. 99).
- ⁴⁹ Ibid., chap. 14, ¶1 (p. 79).
- ⁵⁰ Ibid., chap. 17, ¶2 (p. 106)
- ⁵¹ See Lloyd, *Morality in the Philosophy of Thomas Hobbes*, p. 21; Ewin, *Virtue and Rights*, pp. 3, 66-8.

- ⁵² Ibid., chap. 5, ¶3 (p. 23). Gauthier focuses on this passage in his fascinating essay "Public Reason," *Social Philosophy & Policy*, vol. 12 (1995): 19–42. I discuss his interpretation in my *Contemporary Theories of Liberalism* (London: Sage, 2003), chap. 3.
- ⁵³ Hobbes, *Leviathan*, chap. 26, ¶21 (p. 180).

- ⁵⁶ Hobbes, *Leviathan*, chap. 37, ¶13 (p. 300). This remark occurs in Hobbes's discussion of belief in miracles.
- ⁵⁷ Ewin, *Virtue and Rights*, pp. 32ff. See also my *Justificatory Liberalism* (New York: Oxford University Press, 1996), p. 184.
- ⁵⁸ Hobbes, *Leviathan*, chap. 18, ¶1 (p. 110). Italics in original; caps added.

⁶⁴ Hobbes, *Leviathan*, chap. 29, ¶6 (p. 212). I argue for a more absolutist interpretation of Hobbes — which sharply distinguishes him from Locke, in "Hobbes's Challenge to

⁵⁴ Ibid.

⁵⁵ Ewin, *Virtue and Rights*, p. 67.

⁵⁹Ibid., chap. 7, ¶2 (p. 35).

⁶⁰Ibid, chap. 37, ¶13 (p. 299), and in note 21 on this page to the Latin edition.

⁶¹Ibid., chap. 37, ¶13 (p. 300). Emphasis added.

⁶² Ibid., chap. 32, ¶5 (p. 246); chap. 37, ¶13 (p. 300).

⁶³I have greatly benefited from discussions with Shane Courtland on this matter.

Public Reason Liberalism" in *Hobbes Today*, edited by S.A Lloyd (Cambridge: Cambridge University Press, forthcoming)

- ⁶⁵ See e.g., Hobbes, *Leviathan*, chap. 21, \P 7 (p. 139).
- ⁶⁶ Ibid, chap. 21, ¶7 (p. 138). See Ewin, *Virtue and Rights*, p. 20. There is a deep complication raised by Hobbes's puzzling claim that the civil law and the law of nature contain each other. Ibid., chap. 26, ¶8 (p. 174).
- Hampton is famous for arguing that, although Hobbes sets out to justify an alienation contract, in which citizens alienate all their rights to the sovereign, he ends up with an agency contract along the lines of Locke, in which one is only bound to obey so far as the sovereign is performing the job which he was instituted to perform. See her *Hobbes and the Social Contract Tradition*, chap. 7. See also Gauthier, "Public Reason" and Shane Courtland, "Public Reason and the Hobbesian Dilemma," *Hobbes Studies* Vol. 20 (2008): 63-92.

- In the Latin edition of *Leviathan*, as noted in Curley's edition of *Leviathan*, p. 91, note
 Lloyd stresses this point; *Morality in the Philosophy of Thomas Hobbes*, p. 303.
- ⁷² Lloyd develops a sophisticated account along these lines. *Morality in the Philosophy of Thomas Hobbes*, pp. 296-322. Lloyd calls her account of the reply "deflationary:' Ewin essentially ignores the fool.

⁶⁸ Lloyd, *Morality in the Philosophy of Thomas Hobbes*, p. 23.

⁶⁹ Ewin, Virtue and Rights, pp. 87, 161, 199

⁷⁰*Leviathan*, chap. 20, ¶3 (p. 128).

⁷³ Hobbes, *Leviathan*, chap. 15, ¶1 (p. 89).

⁷⁴ Ewin, *Virtue and Rights*, p. 17. Citation omitted.