

Hobbes and normative egoism

Alex Worsnip

Forthcoming in *Archiv für Geschichte der Philosophie*; please cite published version

Abstract: Is Hobbes a normative egoist? That is: does Hobbes think that an agent's normative reasons are all grounded in her own good? A once-dominant tradition of Hobbes scholarship answers 'yes'. In an important recent work, however, S.A. Lloyd has argued that the answer to the question is 'no', and built an alternative non-egoistic interpretation of Hobbes that stresses reciprocity and mutual justifiability. My aim in this paper is to articulate and defend an original 'middle way' interpretation of Hobbes which steers a course between an excessively egoistic and what we might call an excessively 'moralistic' interpretation. According to the interpretation I defend, our obligations have their source in self-interest in the sense that they are all self-assumed results of covenants, our reasons for making which are solely self-interested. But the obligations that result from such covenants can sometimes require us to act against our self-interest.

Is Hobbes a normative egoist? That is: does Hobbes think that an agent's normative reasons are all grounded in her own good? A once-dominant tradition of Hobbes scholarship answers 'yes'. Within this tradition, the interpretative game has been to show how adherence to Hobbesian morality can be justified on purely egoistic grounds. In an important recent work, however, S.A. Lloyd (2009) has argued that the answer to the question is 'no'. Lloyd builds an alternative non-egoistic interpretation of Hobbes that stresses reciprocity and mutual justifiability.

My aim in this paper is to articulate and defend an original 'middle way' interpretation of Hobbes which steers a course between an excessively egoistic and what we might call an excessively 'moralistic' interpretation.¹ On my view, the answer to our central question depends on what we mean by 'grounded in her own good'. Briefly, the view I attribute to Hobbes is this: our moral obligations are all self-assumed results of covenants, our reasons for making which are solely self-interested. But the obligations that result from such covenants can sometimes require us – and give us reason to – act against our self-interest.² So, such obligations are grounded in an agent's own

I am grateful for comments from and/or discussions with Mark Berger, Shelly Kagan, Jonathan Leader Maynard, Christia Mercer, Johan Olsthoorn, Alex Silverman, Ken Winkler, and especially Steve Darwall. I'm also grateful for comments from two anonymous referees for the *Archiv*. Earlier versions of the paper were presented at the Princeton-Penn-Columbia Graduate Conference in the History of Philosophy and the Society for Early Modern Philosophy at Yale; thanks to both audiences for fruitful questions and discussion.

¹ Thanks to Ken Winkler for this way of putting the point.

² I should note at the outset that my account is neutral on some more foundational questions. Though I say that obligation is created by covenant, there are further questions to be answered about how this is possible in such a way that normativity somehow arises out of covenant. I'll stay neutral on whether Hobbes has an account of this; if he doesn't, it may be a substantive problem with his view. There are also questions about what kind of metaethics (realist, constructivist, etc) could support the idea that obligation is created by covenant. Here, I am skeptical that Hobbes has a

good in the sense that they can be traced back to a covenant which she had only self-interested reasons to make. But that does not mean that the reason for keeping the covenant is (solely) that it is in the agent's self-interest to do so. Rather, the act of covenanting creates an obligation on the part of the agent that binds her irrespective of her self-interest.

In part 1, I describe in more detail existing answers to the question of whether Hobbes is an egoist, and develop my own answer to compare and contrast it with them. In part 2, the lengthiest part of the paper, I argue for my interpretation of Hobbes' text. Finally, in part 3, I briefly defend the bare coherence of Hobbes' picture as I have interpreted it, without undertaking to argue that Hobbes' view as I have interpreted is substantively true. My primary aim is to lay out and defend my own original interpretation rather than to criticize alternatives. But I will try to highlight the ways in which my defense of my view exposes the limitations of alternative interpretations.

1. Laying out the possible interpretations

In this first section, I am concerned simply to lay out the conceptual space that determines different possible interpretations of Hobbes. My substantive arguments and textual support for my view will come later.

Egoism: act- and rule-

Let us begin with egoist interpretations of Hobbes. It is natural to distinguish two quite different interpretations within this camp. According to **act-egoist** interpretations of Hobbes, Hobbes thinks that an agent has reason only to perform those individual acts which promote her own self-interest. This view is represented by, for instance, Thomas Nagel (1959), who writes that

“genuine moral obligation plays no part in *Leviathan* at all...what Hobbes calls moral obligation is based exclusively on considerations of self-interest.” (Nagel 1959: 69)

In Nagel's particular view, Hobbes' normative egoism follows from his purported psychological egoism, according to which, as Nagel puts it, “no man can ever act voluntarily without having as an object his own personal good.” According to Nagel, this “excludes the meaningfulness of any talk about moral obligation”, because “moral obligation is something that plays a part in deliberations...[and] nothing could be called a moral obligation which in principle never conflicted with self-interest” (*ibid.*: 74).

A more sophisticated version of the egoist interpretation is often known, by contrast, as **rule-egoism**. According to this view, Hobbes thinks that an agent has reason to perform those acts which conform to the rules which maximize her self-interest over time. This might mean that she sometimes has reason to perform acts which do not maximize her own self-interest. But the idea is

determinate view, and doubt that he would have thought in terms of anything translatable to these terms, though again that could constitute a limitation of his account.

usually that following the rules which maximize her self-interest over time will be a more effective strategy for maximizing her self-interest than simply trying to do so on a case-by-case basis.³

It's important to see that the contrast between act-egoism and rule-egoism is *not* that the latter takes into account long-run effects of acts, whereas the former does not. There is nothing stopping an act-egoist from taking long-run effects of one's acts into account. Rather, the contrast is that rule-egoism considers whether an act falls under a rule that *if followed over time* – that is, on multiple occasions – would maximize the agent's self-interest. By contrast, act-egoism considers only the effects of the individual act being performed (both in the short- and the long-run).

The rule-egoist interpretation of Hobbes still pretty unambiguously counts Hobbes as a normative egoist given the way we have defined 'normative egoism'. For it says that your reasons to perform an act are all given by the fact that the act falls under a rule which maximizes your own self-interest. And that seems fundamentally to ground your reasons in your own good. As such, the rule-egoist notion of 'moral obligation' is only slightly richer than the act-egoist notion.⁴ The rule-egoist can say that you should follow rules that might in some conventional sense be deemed 'moral', but your fundamental reason to do so will have nothing to do with any idea of what you owe to others; it will be grounded in your own self-interest. Hence, rule-egoist interpreters of Hobbes devote much energy and ingenuity to trying to show that following various such rules is the most rational way to pursue one's own self-interest. So, for example, Gauthier interprets Hobbes as holding that the rational pursuit of one's self-interest involves a sort of 'constrained maximization' (Gauthier 1979: 55), whilst Kavka reads him as adhering to a radically risk-averse conception of rationality (Kavka 1995: 21-30).

Here it may be helpful to note a limitation of the obvious parallel between rule-egoism and rule-utilitarianism. Rule-utilitarianism, as standardly understood, is the view that you should only perform acts which fall under a rule which would maximize utility were *everyone* to follow it. Now, there are not one but *two* important differences between rule-egoism and rule-utilitarianism. The first is that rule-egoism makes reference to an agent's own self-interest, whereas rule-utilitarianism makes reference to the general utility. But the second, equally important, difference is that rule-egoism makes reference to rules followed by an individual agent, whereas rule-utilitarianism makes reference to rules followed by *everyone*. So, consider the following chart of possible views:

		<i>Evaluative focal point</i>		
		Individual acts by individual agent	Rules followed by individual agent	Rules followed by everyone
<i>End to be</i>	Agent's self-interest	I. Perform only those acts which maximize your own self-interest	II. Perform only those acts that fall under rules which, when followed by you over time, maximize your own self-interest	III. Perform only those acts that fall under rules which, when followed by everyone over time, maximize your own self-interest

³ For this interpretation of Hobbes, see Gauthier (1969, 1979), Moore (1971) and Kavka (1986, 1995).

⁴ As rule-egoist interpreters of Hobbes recognize: see, e.g., Gauthier (1969: 97).

<i>maximized</i>	General utility	IV. Perform only those acts which maximize general utility	V. Perform only those acts that fall under rules which, when followed by you over time, maximize general utility	VI. Perform only those acts that fall under rules which, when followed by everyone over time, maximize general utility
------------------	------------------------	---	---	---

Rule-egoism, as we have defined it – and as it is defended as an interpretation of Hobbes by Kavka, Gauthier and others – is represented by principle II.⁵ Rule-utilitarianism, by contrast, as usually understood, is represented by principle VI. So we see that the two views differ along two spectra: the end to be maximized, and the ‘evaluative focal point’:⁶ that is, the thing that is being evaluated with respect to whether it maximizes the end.

The contrast in evaluative focal point matters greatly. If rule-utilitarianism were to be understood as principle V, it would be clear that it expresses a purely utilitarian view. It could be rationalized as follows: maximizing utility is what matters; but following principle V is a more effective way to maximize utility than (trying to) follow principle IV; so it’s principle V that you should follow. That might not be a knock-down argument, but it is cogent enough to count principle V as utilitarian, if it is held on those grounds. Principle VI, by contrast, cannot appeal to similar reasoning. For whereas you are in a position to ensure that you yourself follow a rule over time, you are usually not in a position to ensure that everyone follows a rule over time.⁷ So, whereas following principle V can be reasonably construed as part of an actually utility-maximizing strategy, following principle VI cannot be so construed, given the non-compliance of others. Given that, justifying principle VI requires an appeal to something from outside utilitarianism itself: for example, appealing to the idea that you ‘do your part’ even under conditions of non-compliance, or that you have a responsibility not to perpetuate collective action problems, or similar. But these are plausibly non-utilitarian, deontological ideas. So principle VI does not seem to express a purely utilitarian view.

By analogy, principle II can be understood as a purely egoist view, because it can be reasonably construed as part of an actually self-interest-maximizing strategy. If rule-egoism were understood as principle III, though, it would be much harder to argue that it is a purely egoist view. For following principle III cannot be construed as part of an actually self-interest-maximizing strategy, given that your following (or breaking) some rule does not ensure that other agents will actually follow (or break) it too. Once again, then, justifying principle III requires some kind of appeal to something from outside a purely egoist theory – for example, an appeal to the claim that you ought to ‘play fair’ in pursuit of your own self-interest. So principle III is not a pure egoist view.

⁵ See, e.g., Kavka (1995: 30-4) for clarity that the view defended is expressed by principle II and not principle III.

⁶ This helpful term is from Kagan (2000).

⁷ Of course, your acts do sometimes influence others to act likewise. But in such cases, these influences – and their positive effects – will be part of the consequences of your acts themselves. So such effects are already taken into account by principles IV and V. What principle VI *distinctively* takes into account is what would happen if everyone followed the rule, even when your following it is *not* causing them to do so.

However, to reiterate, ‘rule-egoism’ as we are understanding it is captured by principle II, not principle III. So it *does* count as an egoist view. Any recourse to interpreting Hobbes along the lines of principle III, on the other hand, would be to move him away from a pure egoist view. Still, as one can easily see from the chart, it would not be so going so far as to interpret Hobbes as anything like a rule-utilitarian, of the sort of principle VI: it is still the agent’s own self-interest, rather than the general utility, to which we relate the assessment of the rules.

Lloyd’s ‘reciprocity interpretation’

On Lloyd’s interpretation of Hobbes, “one acts against reason when one does what one would judge another unjustified in doing” (Lloyd 2009: 4). On Lloyd’s view, this is the central principle of Hobbes’ moral philosophy, from which the laws of nature (Hobbes’ particular moral principles) are derived. Lloyd calls this view the ‘reciprocity interpretation’. It is crucial to understand what the notion of reciprocity in play is. We can distinguish what we might call *expected reciprocity* theories from what we might call *hypothetical reciprocity* theories. On an expected reciprocity theory, one acts wrongly in failing to reciprocate adherence to principles which one *expects* others to stick to, as things actually are. On a hypothetical reciprocity theory, by contrast, one acts wrongly in failing to reciprocate adherence to principles which one others *would* stick to, under some particular condition. Such a condition might involve idealization from the actual world, or it may – as in Lloyd’s account – involve an explicitly normative condition, so that we are concerned with the principles that others could not *justifiably* violate.

Since Lloyd’s account builds in this normative condition, it is a hypothetical reciprocity theory, and it is crucial to keep this in mind. Now, it may turn out what whether one would judge another unjustified in doing something depends in some important way on whether certain conditions of actual expected compliance are met. Still, on Lloyd’s interpretations it is the ability to *justify* oneself to others that is normatively fundamental here. This view which could be described as a kind of proto-Rawlsian contractualism, and indeed Lloyd appeals to the test of imagining oneself on the receiving end of an action (*ibid.*: 16-17).⁸

Lloyd’s interpretation of Hobbes is profoundly non-egoistic, in a number of ways. First, traditional egoistic interpretations have it both that our reason for complying with the laws of nature is basically egoistic. Indeed, on many egoistic interpretation, the essential function of the laws of nature is to serve as precepts which explain how to further one’s own self-interest or self-preservation. Lloyd sharply rejects both of these claims (*ibid.*: 98), and holds, by stark contrast, that the normativity of the laws of nature is grounded in the requirement that we justify ourselves to others. This already makes their foundation and normativity fundamentally non-egoistic.

So far this still leaves open the possibility that the laws of nature might not ever actually require an agent to act against her self-interest. But Lloyd also strongly rejects this claim. In fact, she holds – quite radically – that

⁸ Besides the affinity with Rawls, there is an even stronger affinity here with the original views of Rawls’ student Thomas Nagel (see esp. Nagel 1970: 82-84). This is somewhat ironic, since as we just saw Nagel himself advances a purely egoistic interpretation of Hobbes (which he believes to be mistaken).

“the only end reliably served by the Laws of Nature is *the common good, or the good of humanity generally*, and not the preservation or profit of the individual agent who is to follow those laws [...] it is actually the function of securing humanity’s good that ultimately unifies Hobbes’s Laws of Nature.” (*ibid.*: 98; her italics)

This passage makes it very clear that when Lloyd talks of the ‘common good’, she uses this term in a way such that what is in the ‘common good’ is not necessarily always good for each individual: indeed, she explicitly says that observing some of the laws of nature may be bad for some individual agents (*ibid.*: 120-31). So, since on Lloyd’s view the laws of nature enjoin us to promote the common good, or good of humanity, it follows that they can require us to act against our own self-interest. In summary, on Lloyd’s view the laws of nature are egoistic neither in their grounding nor in the content of their instructions. Rather, their ground is reciprocity (understood as mutual justifiability), and they enjoin us to the pursuit of the good of humanity.

The middle way

The reader may wonder whether my middle way will be to interpret Hobbes as holding principle III from the table above, which enjoins an agent to perform only those acts that fall under rules which, when followed by everyone over time, maximize her own self-interest. Though this principle is closer to my interpretation than my interpretation is to the rule-egoist of principle II, or indeed the rule-utilitarian of principle VI, it still does not quite, in my view, capture Hobbes’ view.

An exposition of my interpretation of Hobbes’ view requires some very basic Hobbesian concepts. As is well-known, Hobbes is a social contract theorist who uses the device of a hypothetical state of nature. He imagines humans covenanting to leave this state of nature and form a commonwealth, where a sovereign has the power required to enforce the covenant made. Of course, this is a hypothetical device, since most of us have never experienced the state of nature, nor had the chance to make such a covenant. But Hobbes’ view is that without a state we would be cast back into a state of nature, and he has an exceptionally broad notion of consent which takes consent to be present tacitly through the acceptance of the benefits of commonwealth (even under what we would regard as duress!)

If we understand Hobbes as holding that covenanting is an act that changes the normative landscape in terms of our reasons, then the state of nature can helpfully be viewed as a device not just for understanding what the basis of the state is, but also to reveal what our bare normative reasons are before we have covenanted. The ‘before’ here need not indicate temporal priority, but rather a kind of metaphysical priority: examining our condition in the state of nature reveals what our reasons would be *were* we not yet to have committed ourselves to anything by our agency – by our acts of (tacit) consent.

My interpretation of Hobbes, then, is as follows. In the state of nature, all our reasons are purely egoistic – they are grounded in our own self-interest. However, one of the things which it can be in one’s interest to do in the state of nature is to make a covenant which allows one to leave the state of nature. The only reason to *make* this covenant is self-interested. However, once one has

covenanted and is not living in the state of nature, one is then obligated to keep covenant, and to live by certain limited principles of fair play in the pursuit of one's self-interest. And this obligation applies to one regardless of one's self-interest; indeed, it can compel one to act against one's self-interest on individual occasions.

This obligation is fundamentally dependent on having covenanted, which is – again – something that one has only self-interested reason to do. So there is a sense in which all of an agent's reasons are grounded in her own self-interest, since the obligations are assumed for self-interested reasons. That said, the act of covenanting does create an obligation which is irreducible to self-interest in another sense, namely that the obligation to act cannot itself be explained in terms of how its *performance* conduces to the agent's self-interest, even as part of a strategy over time.

This view does have a close affinity with principle III, for the following reason. The rules that one has self-interested reason to covenant into are just those rules which principle III enjoins one to follow. This is because the rules which principle III enjoins one to follow are those which would promote the agent's self-interest if everyone followed them. Since the agent covenants for self-interested reasons, then, and since in a commonwealth there is enforcement of the covenant by the sovereign such that one can expect compliance from one's fellow citizens, the agent has reason to covenant into any rule the general observance of which would promote the agent's self-interest; that is, the rules that principle III enjoins the agent to follow. And if the agent does covenant into those rules, she will be obligated to obey them; and that would be extensionally equivalent to an enjoinder to obey principle III under those circumstances.

However, this does not quite make the middle way view identical to that of principle III. For the enjoinder to obey principle III is still conditional on one having covenanted, and principle III itself makes no reference to such conditionality. In the state of nature, one has no obligation to follow the rules the general observance of which would promote one's self-interest – one may simply pursue one's self-interest. It's only after the self-interested act of covenanting that the obligation to follow them applies; principle III makes no reference to the need for such an initial self-interested act. Moreover, the same goes for any rule the general observance of which would promote one's self-interest, but which others have not actually covenanted into (perhaps because its general observance would not promote *their* self-interest). If there has not been a general covenant into such a rule, then one is not obligated to follow it. Again, principle III makes no reference to this. So principle III and the middle way interpretation are not quite the same.

Lloyd's 'reciprocity interpretation', on the other hand, differs from both principle III and the middle way interpretation in that, on those latter views, there is at most an obligation to follow the rules the general observance of which promote and sustain the agent's self-interest, and not any obligation to follow rules which simply promote the common (general) good, as there is on the reciprocity interpretation. In addition to this, however, pretty much all of these differences between principle III and the middle way interpretation also obtain between the middle way view and the reciprocity interpretation. It's certainly true that the middle way interpretation involves a kind of reciprocity: having covenanted, one is bound to stick by the rules of the covenant – the rules the general observance of which sustains one's self-interest. But again, this is conditional on one's

having covenanted, something which one only has self-interested reason to do; here Lloyd would disagree.

The underlying difference here is that on the middle way view, one is only obligated in cases where one can genuinely expect reciprocity from others in the actual circumstances, not to act in ways which would further the common good on an idealized assumption of reciprocity. In my view, this is a central, crucial difference between Hobbes' moral theory and that of Rawls or Kant.⁹ Lloyd's principles of reciprocity are different. She admits that our obligations are less demanding in conditions of non-compliance – that is, in the state of nature, and in particular when our lives are in danger. However, she denies that this means that considerations of reciprocity and mutual justifiability do not apply in full force in such circumstances – it's just that the content of what they require in such circumstances is less demanding (Lloyd 2009: 116). This difference is subtle but important, and I will return to it when surveying the textual evidence for the different views.

It should also be clear how the middle way interpretation differs from rule-egoism.¹⁰ The idea is not that one should live by the principles of fair play because doing so will be in one's self-interest over the long-run. On the view I am offering, there could in principle be some rule which it would be in one's self-interest *not* to follow, but which one is obligated to follow nevertheless – if one benefits from *others* following it, and thus tacitly covenants into it. One's reason to do so is that one's covenanting creates an obligation that holds irrespective of self-interest; the notion of moral obligation in the middle way view is thus significantly richer than it can be on the rule-egoist view.¹¹ And, of course, it also differs from the simpler and barer act-egoist view.

The middle way view is named for its attempt to steer a middle course between an egoistic and a full-blown moralistic interpretation of Hobbes; but there is one sense in which it sets itself apart from *all* the other views we have surveyed and thus involves an element of radicalism that they are united in lacking. Specifically, the middle way view takes it that the normative landscape is

⁹ More precisely: Hobbes' moral philosophy is very different from Kant's ethics, but similar to Kant's theory of the right, which invokes a similar assurance constraint on obligations of right like those contractual obligations and obligations to respect property rights (MS: 6/307).

¹⁰ David Gauthier, a rule-egoist, actually states something very similar to my view at one point: "Hobbes supposes that all obligations are undertaken on prudential grounds – as means to self-preservation. But this does not make obligation a prudential concept. To say that obligations are undertaken for prudential reasons is not to say that we are under obligation only so long as we find it to our interest" (Gauthier 1969: 60). But then he continues: "But there is, it must be admitted, a difficulty here for Hobbes. We may suppose it possible to oblige ourselves on prudential grounds without taking obligation to be itself prudential, because we suppose that we may have non-prudential reasons for fulfilling our obligations... Now Hobbes does not suppose that we have non-prudential reasons for acting. Practical reason is linked to self-preservation." Given that the performance of obligation can never be against reason for Hobbes, this second passage requires Gauthier to reject the first. If we can't have non-prudential reasons to ϕ , then we can't be under obligation to ϕ even when ϕ -ing is against our prudential interests. Indeed, Gauthier eventually says that there is no such thing as moral obligation proper for Hobbes (*ibid.*: 97), abandoning the view similar to mine. As should be clear by now, I disagree with Gauthier that Hobbes does not allow non-prudential reasons, whilst allowing that such non-prudential reasons are self-assumed for prudential reasons.

¹¹ There is a certain sense in which the middle way interpretation and the rule-egoist interpretation are on the same page. I think that on Hobbes' view self-interest is what gives one reason to engage in a certain set of practices which are constitutive of social behaviour. Internal to these practices, one does not reason in an egoistic manner. On a certain view of rules associated with the early Rawls (1955), that is what is involved in accepting a rule. So a certain kind of rule-egoist might be with me so far. (Thanks to Sonny Elizondo here.) But to keep this account distinctively egoist, I think we have to view this way of thinking within a practice as a kind of heuristic fiction. Otherwise, we have allowed genuinely non-egoistic reasons to arise. By contrast, the middle way interpretation does not think of the practice as involving a fiction.

fundamentally altered in a commonwealth compared with the state of nature. Each of the other views tries to offer a central principle that applies in both states – and to explain the difference in how we should act simply in terms of differences in whether particular background empirical conditions obtain. By contrast, the middle way view takes it that by covenanting, we have the power to transform the normative situation we find ourselves in, from one in which we have only self-interested reasons, to one in which we have moral obligations proper. It thus offers a fundamentally bifurcated account, attempting neither to deflate the moral obligation of the commonwealth into the bare pursuit of self-interest, nor to inflate the self-interested activity of the state of nature into a pseudo-obligation. It simply accepts that we have no obligations proper in the state of nature, but do once we have covenanted.

2. Arguing for the middle way interpretation

I now proceed to argue for the middle way interpretation of Hobbes' view. I do so by proceeding in stages, in a way that emphasizes the bifurcation in Hobbes' account that I have just alluded to. Establishing the bifurcation is crucial to demonstrating the advantages of the middle way account over both the egoist view and the reciprocity interpretation.

In the first stage, I will examine the state of nature, pre-covenant. I will argue that on Hobbes' view, an agent has no obligations whatsoever that could cut against her own self-interest, and thus no moral obligations proper, in this state. Against the reciprocity interpretation, I hold that considerations of mutual justifiability simply do not apply in such a state; it is not merely that their content is less demanding. In the second stage, I will deal with the precept of peace, by which we are enjoined to leave the state of nature. On the middle way view, this act of leaving the state of nature is what puts us into a position where we subsequently are subject to moral obligations. But I will argue that one's reasons to obey the precept of peace, and thus to covenant, are themselves all self-interest. In the third stage, I will deal with the commonwealth, where subjects have covenanted. I will argue, against the egoist view, that one can have genuine moral obligations in this stage, ones that are not reducible to self-interest – even though they are created by the self-interested act of covenanting. But I will also argue that even in a commonwealth, our obligations are not as extensive as the reciprocity interpretation holds – in the ways that I have already indicated in part 1. Throughout, I continue to compare and contrast my interpretation with the other views, and to argue that it makes better sense of the text.

The basic puzzle in discerning Hobbes' position in relation to egoism is as follows. On one hand, his treatment of the foundations of obligation seems quite egoistic, at least on the surface. Central to Hobbes' account is the idea of a 'law of nature', which he famously defines as "a precept or general rule, found out by reason, by which a man is forbidden to do that which is destructive of his life or taketh away the means of preserving the same, and to omit that by which he thinketh it may be best preserved" (L:xiv/3). The laws of nature forbid me to destroy *my* life; specific laws which are other-regarding in content must still be laws only because of their effect on *me*. Yet on the other hand, Hobbes seems to take the content of our moral obligations to often require us to violate our self-interest. He criticizes agents for following their own self-interest and being ignorant of

genuine moral right and wrong (L:xi/21), tells us that the laws of nature are “contrary to our natural passions, that carry us to partiality” (L:xvii/2)¹², and appeals to a kind of what-if-everyone-did-that logic (L:xv/35). My hope is that my interpretation does the best job of squaring these two, seemingly opposing, currents in Hobbes’ text.

Stage one: the state of nature

As I have said, I am going to first argue that in the state of nature, there are no obligations that cut against the agent’s self-interest. Here are two key passages from Hobbes which indicate this:

“The desires and other passions of man are in themselves no sin. No more are the actions that proceed from those passions, till they know a law that forbids them – which till laws be made they cannot know. Nor can any law be made, till they have agreed upon the person that shall make it.” (L:xiii/10).

And again, speaking of the state of nature:

“To this war[re] of man against every man, this also is consequent: that nothing can be unjust. The notions of right and wrong, justice and injustice, have there no place. Where there is no common power, there is no law; where no law, no injustice. Force and fraud are in war[re] the two cardinal virtues.” (L:xiii/13)

Notice that the claim is not just, as Lloyd is willing to concede, that our obligations do not require us to sacrifice our ability to defend ourselves. Rather, it’s that the very notion of an action being forbidden does not make sense until there is a law. In this context, Hobbes must mean a civil law legislated by a sovereign, since he says in the first passage above that people need to agree upon the person who will make it; that wouldn’t be the case with divine or natural legislation.¹³ Furthermore, the point is not one about specific moral concepts such as obligation or justice, but about all notions of right and wrong.

Now, it might be objected here that for Hobbes, even the words ‘right’ and ‘wrong’ signify something narrowly concerned with justice, and not the whole domain of morality.¹⁴ It’s hard to adjudicate on this point, since Hobbes rarely actually uses the term ‘morality’. But what seems clear is that however the sphere of normative reasons beyond right and wrong is to be characterized for Hobbes, it will not involve any kind of duty or obligation that cuts against one’s self-interest. This is clear because anything one does that violates a duty to *another* is classified for Hobbes as a ‘wrong’ (DCv: iii/3). Furthermore, the fact that Hobbes says above that the actions proceeding from passion are no *sin* in the state of nature suggests that in any recognizable sense of moral duty or obligation, there are none in the state of nature. Hobbes even calls force and fraud the cardinal *virtues* of war[re],

¹² See also L:xv/31.

¹³ See also L:xv/2-3.

¹⁴ Thanks to Johan Olsthoorn and Julia Jorati for pushing this objection. See also the interpretation of Hobbes’ account of morality in Ludwig 2004, which appeals to Hobbes’ remarks at L:xxvi/22.

implying that the virtues of the situation we find ourselves in are those that conduce to our self-interest. Those virtues may be moral ones in an ordered society, but they are force and fraud in war[re]. As he puts it in *De Cive*, “in the state of nature the measure of right is interest” (DCv: i/10).

All of this constitutes, I think, compelling evidence contra Lloyd that Hobbes thinks that we do not have any moral obligations in the state of nature. However, there does appear also to be evidence cutting in the opposite direction. First and foremost, there is Hobbes’ contention that laws of nature are “immutable and eternal” (L:xv/38)? How are we to explain this claim, which seems straightforwardly inconsistent with the passages cited above? We can begin by noting a distinction that Hobbes makes:

“The laws of nature oblige *in foro interno*, that is to say, they bind to a desire they should take place; but *in foro externo*, that is, to the putting them in act, not always. For he that should be modest and tractable, and performs all he promises, in such time and place where no man else should do so, should but make himself a prey to others, and procure his own certain ruin.” (L:xv/36)

What does this mean? Lloyd, who naturally wants to stress the immutability of our obligations, claims that the contrast concerns only to whom one is obligated: in the *in foro interno* case, conscience; in the *in foro externo* case, some kind of external authority. Correspondingly, she draws the inference that I am still obligated in the state of nature, and in a way that can cut against my own good – but only to my conscience (Lloyd 2009: 117). This way of reading the distinction, however, does nothing to resolve the tension with Hobbes’ claim that the actions proceeding from passions are no sin until there is a civil law. Furthermore, given that Hobbes thinks we ought to avoid our own self-destruction, it is unclear why the fact that acting on the laws of nature would “procure [our] own certain ruin” would not let us off the hook even from an obligation to act owed only to conscience. If one ought to avoid one’s ruin, why feel guilty about doing so?

So we should look for a different interpretation of Hobbes’ distinction, closer to his own gloss, namely that when laws oblige *in foro interno* “they bind to a desire they should take place.” This seems to indicate that I ought to wish for – and perhaps try to bring about – an order in which the laws of nature are and can be obeyed by everyone, including me – that is, a civil order, which would after all be in my interest. That is not, of course, to say that I am bound to obey the laws now, when everyone else is not complying (see DCv: iii/27). In a limited range of cases, though, I may be able to observe the laws in the state of nature without any relevant fear of reprisal. For example, Hobbes notes that pointless cruelty serves no end of the agent in the state of nature, and thus ought to be refrained from, as it frustrates any hope of peace. But he is very clear that as soon as the agent’s good is under threat in any way (and, the agent himself is to be the arbiter of when this is the case), he no longer need observe the law. This is why an *in foro interno* obligation is strictly limited and not much like an ‘obligation’ in any standard sense. Most importantly, for our purposes, it does not, in the state of nature, involve any enjoinder to the agent which cuts against his own self-interest.

What would cut against self-interest in the state of nature would be an obligation *in foro externo*, which is really what we ordinarily mean by an obligation *simpliciter*: namely, an obligation “to

the putting them in act” – to actually comply with the law in question, come what may. This, I think, is the robust sense of obligation that Hobbes denies can obtain in the state of nature. The contrast between *in foro interno* and *in foro externo* obligation thus concerns not merely *whom* one is obligated to, but what one’s obligation entails. The resulting view is that laws of nature are eternal, but only obligate – in the standard, *in foro externo* sense – in a civil order. This renders the contentions about right and wrong having no application in the state of nature consistent with the comments about the status of the laws of nature.

In fact, Hobbes himself puts the point very clearly: “the laws of nature...in the condition of mere nature...are not properly laws, but qualities that dispose men to peace and to obedience. When a commonwealth is once settled, then are they actually laws, and not before” (L:xxvi/8). So the normative function of the laws of nature changes between the two states, again emphasizing the bifurcation of Hobbes’ account, something which the middle way interpretation stresses.

Now, it is clear that Hobbes thinks that agents have some kind of prudential reason to seek peace in the state of nature. In *De Cive* ii/1, Hobbes tells us that men exercise right reason in their natural state by engaging in their own reasoning about what conduces to their benefit and loss. But this right reason cannot yield genuine duties, apart from anything else because in the state of nature the standard of reason is at the discretion of the agent himself (see DCv: i/9).¹⁵ Interestingly, this seems to imply that one is not even obliged (say, by God) to seek one’s own self-interest according to a prior, independent standard of self-interest. Certainly, in any case, one is not in the state of nature under any duty to pursue something that cuts against one’s self-interest.

Consequently, though “injustice, ingratitude, arrogance...and the rest, can never be made lawful” (L:xv/38), we have the *right* to engage in such behavior in the state of nature, since for Hobbes right is merely absence of obligation to do otherwise.¹⁶ This right is nothing like a claim-right, nor a Kantian ‘title to coerce’, for since right is just absence of obligation to do otherwise, nothing about a right on my part implies a correlative obligation on yours. In the state of war, we both have the right to destroy each other, and the right to defend ourselves. Indeed, “every man has a right to everything, even to one another’s body” (L:xiv/4). If our opposing rights were claim-rights, they would be inconsistent with one another, but since they are not, they bring about practical conflict but involve no conceptual inconsistency at all.

Stage two: leaving the state of nature

It might be thought that I’ve missed a crucial law which applies in the state of nature, namely to seek peace. This precept must apply antecedent to commonwealth, because we need to follow it to get to

¹⁵ This result is not just achieved by substituting an inflationary, modern notion of a duty for a deflationary, Hobbesian one. For, as we saw, it is Hobbes himself who tells us that the laws of nature are not properly laws in the state of nature, and are proper laws in the commonwealth. He thus clearly thinks that the laws have a certain binding force in the commonwealth which they do not have in the state of nature. The word ‘duty’ comes in for Hobbes only in dealing with what the dictates of prudence *draw us towards*, not in characterizing the dictates of prudence themselves (see DCv: ii/1). We will see in section 3 that Hobbes does allow for a more robust kind of duty or obligation in the commonwealth, which reinforces the point that the dictates of prudence in the state of nature are not as robust as Hobbesian normative concepts get.

¹⁶ See L:xiv/1, 3; xxvi/8.

a commonwealth in the first place. But I'll now argue that, until we have a commonwealth, this precept enjoins an agent to action only insofar as it advances the good of that agent. Thus, it does not conflict with the middle way interpretation, which claims that until the commonwealth, the laws of nature do not issue in obligations that cut against an agent's self-interest.

Hobbes thinks the setting up of the sovereign power is in everyone's private interest (L:xv/40), and that this is what motivates them to seek peace (L:xvii/1). Consequently, it is also in the common interest, as a sum of private interests. But because leaving the state of nature really is in *each* individual person's interest, the obligation to seek peace – unlike that to keep covenant – does not involve any action that is even *prima facie* non-egoistic. Still, it is important to ask what it is that actually normatively grounds agents' reason to seek peace in the state of nature: private interest, or common interest? Here is Hobbes:

“As long as this natural right of every man to everything endureth, there can be no security to any man (how strong or wise soever he may be)...and consequently it is a precept, or general rule of reason that every man ought to endeavour peace, as far as he has hope of obtaining it, and when he cannot obtain it, that he may seek and use all helps and advantages of war.” (L:xiv/4)

In several ways, this passage indicates that it is our private interest, rather than the common interest, that enjoins us to leave the state of nature. First, note the passage's placement: it comes straight after the definition of a law of nature as “a precept or general rule, found out by reason, by which a man is forbidden to do that which is destructive of his life”, and before the concepts of justice, reciprocity and covenant-keeping have been introduced. Thus, when Hobbes tells us that it is a “precept, or general rule that every man ought to endeavour peace”, we have no reason to take this to be a rule pertaining to anything but the agent's own good. Justice and reciprocity are to be explained in terms of how they secure peace, not the reverse.

Furthermore, the fact that Hobbes feels the need to point out that in the natural condition there can be no security to *any* man, even the strongest, implies that he feels he needs to show that peace is in *everyone's* interest to ground the rule of reason by which all agents are enjoined to seek peace. If agents had reason to seek peace on account of the general good, we would only need the point that peace will generally improve society's overall well-being.

Last, each individual agent is only enjoined to endeavour peace insofar as he has hope of obtaining it. Consequently, we are not yet enjoined to a universalized ‘what-if-everyone-did-that’ form of moral reasoning. We cannot be bound by a non-egoistic moral obligation in the state of nature, because then we wouldn't have a ‘right to everything’. But when reason enjoins us to seek peace for our own advantage, it does not restrict natural liberty. Only an egoistic rendering of the precept to seek to peace allows us to square it with the right to everything in the state of war.

The only way to provide the peace which all individual men must rationally seek is to enter a commonwealth, erecting a sovereign power to enforce civil laws. But entering this commonwealth requires us to lay down some of our natural liberty.¹⁷ Hence the second law of nature: “that a man

¹⁷ See DCv: preface/14 for a particularly clear statement.

be willing, when others are so too, as far-forth as for peace and defence of himself he shall think it necessary, to lay down this right to all things, and be contented with so much liberty against other men, as he would allow other men against himself” (L:xiv/4). The individual agent’s reason for laying down his right is egoistic. However, Hobbes thinks that the only way to guarantee peace is to have a system of parity whereby men lay down their rights in equal proportion: A cannot claim any rights against B which A would not allow B to possess against A.

In section 3 I will argue that this system, once in place, forbids one to break covenant for one’s own advantage. But here the important point is that the reason to *enter* the system is based solely on egoistic considerations. For the laws of nature which Hobbes derives from the fundamental precept of peace are those which, when established as general rules, further the egoistic interest of each individual. Hobbes indicates this when he writes that

“The passions that incline men to peace are fear of death, desire of such things as are necessary to commodious living, and a hope by their industry to obtain them. And reason suggesteth convenient articles of peace, upon which men may be drawn to agreement. These articles are they which otherwise are called the laws of nature.” (L:xiii/14)

The laws of nature function as ways of furthering the individual passions of men, not a transcendent common good. Reason does not enjoin us to overrule our passions (at least, not our stable, far-sighted ones), but it does suggest ways in which we can realize our passions better than if we all follow them in an uncoordinated manner. Hence Hobbes’ focus on mutual agreement. If we all agree upon a set of binding rules, we can realize our passions better. None of this commits Hobbes to the view that in every instance, following a law of nature is to the agent’s benefit. It does, however, commit him to saying that the gain the agent gets from it being a precept which everyone follows well outweighs any individual losses she might incur, thus giving her egoistic reason to contract in.

Still, were she not to contract in, she would not be bound by the laws of nature because she would not be part of the social system. We can see this when Hobbes argues that minorities who dissent still consent due to their membership of the congregation:

“Because the major part hath by consenting voices declared a sovereign, he that dissented must now consent with the rest...For if he voluntarily entered into the congregation of them that were assembled, he sufficiently declared thereby his will (and therefore tacitly covenanted) to stand to what the major part should ordain...He must either submit to [the congregation’s] decrees or be left in the condition of war he was in before, wherein he might without injustice be destroyed by any man whatsoever.” (L:xviii/5)

We might be unimpressed by Hobbes’ argument that minorities who dissent can still be taken as having consented due to their membership of the congregation. But in feeling the need to make it,

he indicates that he thinks consent crucial.¹⁸ Indeed, he admits that dissenters can refuse to submit – thereby opting out not only of the laws, but the whole congregation of society. Then, they cannot expect the benefits of the state’s protection. So rationally, they ought to opt in, but if they do not, they are not bound by the civil laws.

Put this together with Hobbes’ view that natural laws bind only insofar as they are legislated as civil laws, and you get the view that we are obliged by natural laws as a result of covenant. Hence, Hobbes says that the reason that there is no law over “natural fools, children, or madmen” is that they “never had power to make any covenant” (L:xxvi/12). Again, we see that the law’s binding force comes solely from covenant. Thus, I believe that Lloyd is mistaken when she says that the laws of nature are duties, not obligations, where only the latter depend upon consent or covenant (Lloyd 2009: 49). Insofar as Hobbes recognises any kind of normative reason at all in the state of nature, such reasons do not amount to duties in anything like the robust sense that Lloyd wants, as I have already argued in stage one. On my view, Hobbes has no distinction between duties and obligations, because the whole system of laws is covenanted into: that is, it is only as a result of the broad covenant to enter the congregation of society that any of the laws of nature have fully binding status. That is why the very definition of wrong or injustice is the breaking of covenant.¹⁹

This also explains why we aren’t obligated in the state of nature. For Hobbes thinks that covenants are void when there is no assurance of compliance, and the only way to assure mutual compliance is to erect a sovereign power.²⁰ Furthermore, Hobbes says that even in the commonwealth, if the sovereign prescribes no rule, the subject always has the right to act as he wishes. Again, this suggests that there needs to be positive agreement, embodied in the sovereign’s prescriptions, for natural laws to bind.²¹

This applies even to the precept of peace, which is itself the first law of nature. We have self-interested reason to seek peace in the state of nature. But the obligation to seek peace only becomes a morally binding one, which we must observe irrespective of self-interest, once we have covenanted. The first law of nature is a special case in one respect because if we successfully follow it (egoistically) in the state of nature, we will thereby exit the state of nature and put ourselves into a state where it, and the other laws of nature, become binding. In general, we lack the capacity to make covenants in the state of nature, because they cannot be enforced. The one exception is a covenant to leave the state of nature, which creates the conditions for its own enforcement. By putting us into a commonwealth, the covenant makes the enforcement of compliance both concretely possible and normatively legitimate. There is thus no paradox about how we can covenant out of the state of nature despite having no obligations to perform covenants in it.

I have claimed that for Hobbes, we are only obligated in a civil order that we have covenanted into. However, in one passage Hobbes appears to contradict this:²²

¹⁸ Hobbes’ notion of consent is very permissive, perhaps so as to make these kinds of argument work. For example, he also holds that covenants entered into by fear are just as obligatory as any other (L:xiv/27), and that indentured servants can tacitly consent to their servitude (L:xx/10).

¹⁹ For an especially clear statement, see DCv: iii/3.

²⁰ See L:xiv/18; L:xv/3; xvii/1-5; xviii/4.

²¹ See L:xxi/18.

²² Thanks to Steve Darwall for pushing me on this point.

“For the question is not of promises mutual where there is no security of performance on either side (as when there is no civil power erected over the parties promising), for such promises are no covenants, but *either* where one of the parties has performed already, *or* where there is a power to make him perform, there is the question whether it be against reason, that is, against the benefit of the other to perform or not. And I say it is not against reason.” (L:xv/5; my italics)

Noting the disjunctive ‘or’ here, which I have italicized, this seems to imply that in a case where one’s co-covenanter has already performed, one is obligated to perform even when there is no power to make one do so. As I have shown, Hobbes clearly holds elsewhere that there is no injustice or obligation without a civil law. So when he seems to admit the possibility of an obligation in the state of nature here, a contradiction arises in his account.

No amount of philosophical acrobatics by interpreters of any sympathy can make the claim that there are no obligations in the state of nature consistent with the claim that there is at least one obligation in the state of nature. However, I am going to argue that Hobbes could have made his account consistent without compromising its most fundamental tenets. Two crucial tenets that I have been emphasizing about his account are that first, that all obligation is the result of covenant; second, that without security of performance, covenant is void.

As a general matter, there is no security of performance in the state of nature. However, a case in which one’s co-covenanter has *already* performed is a special one. Here, there is no concern about whether one’s co-covenanter will perform, because she already has. Hence, the insecurity which normally gets one ‘off the hook’ from one’s obligation in the state of nature does not apply here. So Hobbes can allow obligation in such a case without violating the second tenet. This then explains the disjunction in the passage cited. Likewise, this is a case where a specific covenant has been made, and so he can allow obligation without violating the first tenet.

This view would go like this: say two people, A and B, make a state-of-nature agreement at t1. As it stands, the agreement is not enforceable, and so neither of them are obligated to comply with it. However, say that A complies with it anyway. At t2, after A has complied, B is now obligated to do his part in a way that neither A nor B were at t1. A wasn’t obligated at t1 because she had no guarantee that B would hold up his end of the bargain. B is obligated at t2 because A already had held up her end of the bargain.

At no point does Hobbes say that one is bound by agreements generally in the state of nature. All he says is that one is obligated when one’s co-covenanter has already performed. It is only in a civil order that A and B will be obligated at t1, before either has performed. All this is explained neatly by the account I have just proposed.²³ Our obligations in the state of nature would remain special cases, since they only arise in cases where A performs at t1 despite not being obligated and having no security of reciprocation. This does require Hobbes to qualify his claim that

²³ This account seems quite similar to that offered explicitly in the earlier *De Cive*, where Hobbes writes that “in the state of nature agreements made by a contract...are invalid if a just cause for fear arises on either side”, but goes on to make clear that normally in the state of nature such a just cause for fear is present (DCv: ii/11-13).

there are no obligations without civil law, but not to revise the underlying views which led him to make that claim: that obligation requires covenant, and covenant requires security of performance.

Alternatively, of course, we could just say that Hobbes slipped up in this passage, and didn't mean to acknowledge the possibility of an obligation without an enforcing civil power. This might be supported by the bracketed section of the passage in question, which seems to reiterate Hobbes' view that when there is no civil order, there is no security of performance, and therefore no obligation. Very tentatively, we might speculate that Hobbes realized his mistake by the time of the Latin edition, in which he asks whether it is against reason to perform "if there is a compelling power *and* if the one party has performed his promise" (my italics; see xv/5 Curley fn. 5). Either way, the take-home point remains the same: no obligation without covenant, our reasons for making which are self-interested.

Stage three: the commonwealth

I'll now argue that once we *have* opted into the commonwealth, we have genuine obligations which can cut against self-interest. This can be cashed out in terms of a limited kind of fair play: if I want to gain from the system of cooperation, I have to do my agreed part to uphold it. I will argue that for Hobbes, this ends up being equivalent to an obligation to observe the laws of nature, since we (tacitly) covenant into an agreement to obey precisely those laws.

As I have already argued, Hobbes thinks that in order to leave the state of nature we must lay down our rights. We create civil laws by which "the natural liberty of man...[is] abridged and restrained...in such a manner as [men] might not hurt, but assist one another" (L:xxvi/8). Crucially, although we lay down our rights for egoistic reasons, in doing so we assume (non-egoistic) obligations to our co-covenanters. As Hobbes puts it,

"When a man hath in either manner abandoned or granted away his right, then is he said to be obliged or bound not to hinder those to whom such right is granted or abandoned from the benefit of it...and that such hindrance is injustice." (L:xiv/7)²⁴

As the subsequent paragraphs make clear, this laying down of one's right is done for egoistic reasons. This establishes the limits of which rights one can alienate, since the laying down of some rights (such as that to resist assault by force) could not be expected to be to the agent's own good. Nevertheless, the obligation that one takes on by laying down one's right – indeed, the idea of laying down one's right itself – goes beyond a simple egoism, such that egoist interpretations of Hobbes (whether act- or rule-) cannot make sense of it.²⁵ According to such interpretations, the obligation to observe the laws of nature hangs upon the fact that violation would be against one's self-interest, either on individual occasions or over the long-run. But such an 'obligation' would involve no 'laying

²⁴ See also L:xiv/17; xxvi/41, 44; also Gauthier 1969: 40-3.

²⁵ Some egoist interpreters (especially Nagel 1959, but also in some ways Gauthier 1979) argue to the contrary that there is no room for genuine obligation in Hobbes' conceptual scheme because Hobbes' psychology only allows an agent to act for self-interested reasons. I believe that this view has already been convincingly refuted elsewhere; see Gert 1967 and especially Darwall 2000.

down' of right or natural liberty. It would just involve using one's natural liberty and rights *prudently*. And, as Hobbes makes clear, obligatory law is not mere prudential advice or counsel.²⁶ The idea that one has 'laid down' one's natural liberty implies that one can no longer rightfully use it, not just that, as a matter of fact, one will not do so if one acts in accordance with prudence or even reason.²⁷

Indeed, it follows for Hobbes from the *concept* of covenanting that one is obligated to keep one's covenant, for unless "men perform their covenants made...covenants are in vain, but empty words, and...we are still in the condition of war[re]" (L:xv/1). If one weren't genuinely obligated to keep one's covenant, it wouldn't be a covenant, but empty words. This is very clear in a passage from *De Cive*:

"In making an agreement, one denies by the very act of agreeing that the act is meaningless. And it is against reason knowingly to take away the meaning of anything. If he does not believe the agreement should be fulfilled, by the very fact that he so believes he affirms that the agreement is meaningless...Therefore either one should keep faith with every one or one should not make agreements." (DCv: iii/2)

Our self-interested reason to covenant is thus one to create an obligation which is not itself excused by self-interest. It might be asked here why our self-interested reason is to actually covenant, rather than to pretend to covenant.²⁸ But remember that Hobbes thinks that being part of the congregation, and accepting the benefits of the state's protection, is to tacitly consent.²⁹ Thus, there is no such option as 'pretending to consent' which allows one to reap the benefits of this pretend-consent. Reaping the benefits is consenting.

An advocate of an egoist interpretation of Hobbes may now object that, if obligations in the commonwealth can cut against our self-interest, we cannot explain Hobbes' contention that laws of nature, including the enjoinder to perform covenants, are rules by which a man is forbidden to do that which is destructive of his life. To begin to answer this objection,³⁰ notice that Hobbes' view cannot be that any infraction of law is bound to lead to the agent's destruction. Rather, the argument is that to break the law is to exhibit a pattern of behavior which, if it is reciprocated by others, leads to war[re]. Consider how the laws are presented in the introduction to *De Cive*:

"All men, by necessity of their nature, want to get out of that miserable and hateful state [of nature], as soon as they recognize its misery. But they can only do so by entering into

²⁶ See, e.g., DCv: xiv/1.

²⁷ See also DCv: ii/3-4, 10.

²⁸ Thanks to Mark Berger for pushing me on this point.

²⁹ See, e.g., DCv: ii/16.

³⁰ By contrast with the middle way account, Lloyd's account *is* subject to this objection. Lloyd claims that when Hobbes says that a law of nature is that "by which" a person is forbidden to do that which is destructive of his own life, this need not mean that this is what it is to be a law of nature, but may merely mark a property which laws of nature have, namely, of tending against the agent's destruction (*ibid.*: 107-8). It seems to me rather odd to mark a non-essential property in a definition, even with a "by which" locution. But even if we accept that definitions can mark non-essential properties, Lloyd's interpretation does not take the laws of nature as necessarily promoting *each* individual agent's good, as mine does. So even if we accept Lloyd's nature the laws of nature do not *essentially* tend toward the prevention of the agent's destruction, she cannot explain why the laws of nature are *coextensive* with the prevention of the agent's destruction.

agreements to give up their right to all things. I then proceed to explain and confirm what the nature of agreements is, how rights must be transferred from one to another for valid agreements to take place; likewise what rights must necessarily be ceded to establish peace and to whom, i.e. what are the dictates of reason which may properly be called natural laws.” (DCv: preface/14)

This is a pretty clear statement of the view I have been attributing to Hobbes. It makes clear that people enter into covenants out of self-interest, but that in doing so they lay down some of their rights. The resultant system of obligations is given by “the dictates of reason which may properly be called natural laws”. These are the laws required to establish peace, and thus to prevent the destruction of agents. But it is their general observance which prevents the agent’s destruction. A restriction of an agent’s natural right couldn’t possibly be in his interest *qua* restriction of *his* right – after all, natural rights allow the agent total discretion to pursue the optimal self-interested strategy. But the restriction of his natural right is in his interest *qua* covenant, since it brings him the benefits of others having their rights reciprocally restricted. As Hobbes puts it, the laws of nature are those “duties towards other men [that] are necessary to [the agent’s] own preservation” (DCv: ii/1). The agent must take on the duty in order to preserve himself; that does not imply that every instance of complying with the duty is an act of self-preservation.

This characterization of the laws of nature, then, pushes us away from the interpretation that any violation of a law of nature would destroy the agent, and towards the interpretation that a reciprocated pattern of violation would destroy the agent. We can understand the laws of nature as forbidding behavior which, as Hobbes puts it in his reply to the Foole, ‘tends’ towards the agent’s own destruction (L:xv/5).³¹ In this passage, Hobbes is clear that this quality of ‘tending’ to the agent’s own destruction can persist even in actions which do not actually destroy the agent.³²

Now, on traditional egoist interpretations, an action only ‘tends’ to the agent’s own destruction if the action *will* provoke retaliation from others, or at least is expected to do so. My suggestion, conversely, is that for Hobbes an action ‘tends’ to the agent’s own destruction if it would destroy the agent, *were* others to reciprocate, regardless of the probability of such reciprocation. Why think this? First, note that Hobbes says that someone “seek[s]...the destruction of his nature by violence” by breaking laws even when he has “sufficient security that others shall observe the same laws towards him” (L:xv/36). So even when I cannot rationally *expect* retaliation, my action nevertheless in Hobbes’ sense ‘seeks’ my destruction when *were* others to retaliate reciprocally, I would be destroyed, and it seems reasonable to take this as amounting to it *tending* to my destruction also.³³ Consider also the following passage:

³¹ Hobbes also uses this locution in the Latin version of his formulation of the definition of a Law of Nature: see L:xiv/3, Curley fn. 3.

³² “When a man doth a thing which...tendeth to his own destruction (howsoever some accident which he could not expect, arriving, may turn it to his benefit), yet such events do make it reasonably or wisely done.” (L:xv/5)

³³ The egoist might here try to concede this and argue that for Hobbes, the mere fact that an action *tends* to my destruction makes it against my self-interest to perform (thanks to Johan Olsthoorn for this ingenious suggestion). But, combined with the view of what it is to ‘tend’ to my destruction that I have offered here, this leads to the paradoxical claim that it can be against one’s self-interest to perform an action which is not even rationally expected to have negative consequences for one’s own good. That, I think, is too much to bear. Such an ‘egoism’ would be egoism in name only, and not in any recognizable sense of ‘self-interest’. Even if Hobbes’ had the required, highly non-standard, notion of

“The obligation a man may sometimes have, upon the command of the sovereign, to execute any dangerous or dishonorable office, dependeth not on the words of our submission, but on the intention, which is to be understood by the end thereof. When, therefore our refusal to obey frustrates the end for which the sovereignty was ordained, then there is no liberty to refuse; otherwise there is.” (L:xxi/15)

Hobbes holds that there is one case in which the sovereign can compel me to execute a dangerous or dishonorable office: when my refusing to do so would undermine the end for which the sovereignty was ordained. Note that the claim is not merely that it is wise for the sovereign to compel me in such a situation, from his perspective, nor even merely that it is legitimate for him to do so. It is that I am *obligated* to comply, even against my own self-interest, from my perspective: I have no rightful liberty to refuse. And that is not just because whatever the sovereign says goes, or because the threat of sanction is too great, for Hobbes in the same passage indicates that I do have rightful liberty to refuse, against the sovereign’s order, to kill myself, accuse myself or kill my fellows in other cases (L:xxi/12-13).

What makes the difference between the cases where I do and do not have rightful liberty to refuse is that in the latter, there is a covenant that binds me: the reason, for example, why I am not bound to accuse myself is that “no man...can be obliged by covenant to accuse himself” (L:xxi/13). Conversely, in the case where I do not have rightful liberty to refuse, my obligation is a result of my own ‘intention’ in submitting to the sovereign, namely to be protected. That was the end for which the sovereignty was ordained, and it is an end which I shared; consequently, my self-interest instructed me to consent into a system whereby the sovereign has power. In that system, the sovereign can compel me to do things against my own self-interest, but only if the reason for doing so is to uphold the end which guarantees my self-interest.

On one hand, this displays a normative sophistication which speaks against the egoist interpretation. What makes the difference between the case where I can rightfully refuse and the case where I cannot is not whether the action is in my self-interest. After all, since the sovereign’s power is so great, it may be in my self-interest to submit to his command to accuse myself or kill my fellows, yet in that case I have rightful liberty to refuse. Rather, what makes the difference is whether I have committed myself to upholding the end in question by covenant, relying on it for my own protection. This goes far beyond egoist considerations of actual retaliation and displays the kind of fair play reasoning I have been urging.³⁴

self-interest – and I don’t see any reason to think he did – the substantive point that one can be required to perform actions against one’s own good would have been conceded. Furthermore, such an egoism would be without a story as to why it is that I should care about this odd sense of ‘self-interest’, rather than the actual effects of the action on my own good. Conversely, my explanation in terms of fair play explains why I should care about the ‘tendencies’ of my actions and not just their actual results.

³⁴ The rule-egoist is not better off than the act-egoist here. Recall from part 1 that the rule-egoist account considers the effect on agent’s self-interest of *just that agent* following the rule on all occasions, and unlike rule-utilitarians, they cannot appeal to further other-regarding moral notions to explain why exceptions to rules as unwarranted. But as we have seen, it matters to Hobbes whether I *rely on others* observing a rule to sustain my own self-interest. That is why Hobbes can appeal to a kind of ‘what-if-everyone-did-that’ reasoning (see, e.g., L:xv/23).

On the other hand, it also speaks against Lloyd's more ambitious interpretation. For the causes in which I can be asked to violate my self-interest are quite strictly limited. Any obligation on me that requires me to violate my own self-interest must be grounded in a law the general observance of which is required for my protection. It is not enough that my sacrifice promote the common good.³⁵ This all points to my interpretation of the laws of nature, on which they forbid behaviour which, *were* others to reciprocate, would destroy the agent – whether or not others *will* reciprocate. That, indeed, is what it is for an action to 'tend' to my destruction. In breaking the law, one renders one's covenant void and can no longer *legitimately* rely on others keeping to it either.

This links the notion of 'fair play' to the crucial role of civil law in Hobbes' system. The rules which I am obligated to play fair with in a commonwealth – the laws of nature – are identical to each of the following:

1. the rules which I can genuinely expect reciprocal compliance from my fellow citizens
2. the actual civil laws of the commonwealth, which we have each consented to (tacitly or otherwise)
3. the rules that that we all have self-interested reason to covenant into
4. the rules that I rely on others observing for the preservation of my own self-interest

Why do these characterizations turn out to be equivalent? (1) is equivalent to (2) because there is assurance of compliance with a rule just if that rule is enforced by the sovereign, and the sovereign enforced all and only the civil laws. (2) is equivalent to (3) because those laws which we each have reason to consent to are just those which, when collectively followed, protect each agent's self-interest. And (3) is equivalent to (4) because I consent for self-interested reasons just if I determine that the benefits of other complying is worth the burden of complying myself.

This all contrasts with Lloyd's picture, on which the duty to follow the laws of nature is derived, from an *independent* principle of reciprocity, rather than being assumed as a result of covenant entered into from self-interest or dependent on the actual compliance of others.³⁶ On Lloyd's view, we are bound to comply with a rule whenever such compliance is demanded by justification of oneself to others, which turns out to be whenever it provides an opportunity to further the common good (understood such that my own good can conflict with the common good).

On the present interpretation, by contrast, there may be spheres of action in which it makes sense to me to 'go it alone', and pursue my own self-interest at cost to others and even the common good. This is consistent with fair play as long as I am willing to accept non-cooperation from others on the same front. There is no requirement of mutual justifiability here because there has been no covenant.³⁷ Such action is not forbidden by the laws of nature, because the laws of nature concern precisely those areas in which it is not possible for us to go it alone without *relying* on the compliance of others. We covenant into the observance of these laws because we need others to observe them

³⁵ Lloyd's argument that only a common good account can explain how one is obligated to protect the sovereign at cost to one's self-interest (2009: 54) is thus too quick.

³⁶ Hobbes thinks, however, that obligation requires assurance of mutual compliance; see L:xiv/5; DCv: iii/27.

³⁷ See DCv: iii/4.

to pursue our own self-interest. It is only then that fair play kicks in. The idea is not one of full-blown reciprocity requiring the pursuit of the common good,³⁸ but of fair play in the pursuit of our own interests within co-operative schemes that we have voluntarily joined.

There is one more major hurdle for the middle way to clear: the reply to the Foole. Here again the pressure comes from the direction of the egoist interpretation. The Foole alleges that in cases in which breaking a covenant benefits the agent, it is not contrary to reason to do so (L:xv/4)?³⁹ Why doesn't Hobbes just cite the obligation to keep covenants and be done with it? Independently of my interpretation, Hobbes' text clearly makes such a reply possible. For instance, Hobbes endorses the golden rule – he could have used that. So why didn't he?

Let's get clear about the Foole's position. Hobbes is clear that the Foole "does not...deny that there be covenants, and that...breach of them may be called injustice, and observance of them justice; but he questioneth whether injustice...may not sometimes stand with that reason which dictateth to every man his own good" (L:xv/4). So far, the Foole's position makes no claim that one is permitted to break covenants. All it does is question whether the keeping of covenant is always in accordance with self-interest. However, the Foole goes on to say that covenant-breaking in one's own self-interest "can never be against reason, seeing all the voluntary actions of men tend to the benefit of themselves, and those actions are most reasonable that conduce most to their ends." This passage has the Foole going further in claiming that it can be *reasonable* to break covenants, making the substantive claim that acting in one's self-interest is never unreasonable.

Now, this passage is a direct quotation of the Foole, which Hobbes describes as "specious reasoning". However, in setting up the point of contention between himself and Foole, Hobbes says that the question is "whether it be against reason, that is, against the benefit of the [agent] to perform [his covenant] or not" (L:xv/5). This seems like it might concede to the Foole that nothing that concords with one's own self-interest can be unreasonable.

However, when we look elsewhere, we find that Hobbes does not share this picture. Remember that the Foole's position is that covenant-breaking can be unjust, but not against reason if it conduces to the agent's benefit. Kinch Hoekstra (1997) has attributed this position to Hobbes himself. But Hobbes says that justice (which consists in keeping covenants) signifies "conformity or inconformity to reason...of particular actions (L:xv/10), and that "law can never be against reason" (L:xxvi/7).⁴⁰ Indeed, these passages – particularly the word 'signifies' – suggest not even a contingent co-incidence between reason and justice, but a conception of reason whereby it is necessarily unreasonable to act unjustly.

So why does Hobbes drop this conception in the reply to the Foole? I suggest that when Hobbes writes "against reason, that is, against the benefit of the [agent]", he is using "that is" not to explicate a substantive view about what reason is, but to stipulate that this is how he will use 'reason'

³⁸ Hobbes denies the existence of an utmost end at which we ought to aim (L:xi/1) – the precise function which Lloyd takes the common good to play.

³⁹ Thanks to Daniel Garber for criticisms of my interpretation of the reply to the Foole, which prompted me to develop and partially modify my account.

⁴⁰ It might be objected here that Hobbes cannot really believe this, since he allows violation of the laws of nature in the state of nature (see Hoekstra 1997: 638). But remember that Hobbes says that the laws are not properly laws before there is a civil order, and, in particular, that would-be covenants are no covenants at all without enforceability; so, it is never against reason to violate an actual law or covenant.

here. This is really not an unusual thing to do with a word like ‘reason’ or ‘rational’ that has multiple uses: one finds even philosophers with very precise views about rationality acceding to common usage in certain contexts. Since this is a context of dialectical engagement, he switches to the terminology of his interlocutor. Where Hobbes is setting out his own positive view, he does not endorse the Foole’s egoistic conception of reason.

Furthermore, the later Latin edition of the relevant passage is even weaker, with the equation between reason and self-interest dropped entirely and replaced by a conjunction: “the question is then whether the one who deceives does so with reason *and* in accordance with his own good” (my italics). Similarly, note that when Hobbes is presenting the Foole’s position from an observer’s perspective, he says that he “questioneth whether injustice...may not sometimes stand with that reason which dictateth to every man his own good. The use of ‘that’ as a pronoun here suggests that this is only one perspective of reason, not an all-things-considered claim about reasonableness. It is only in the Foole’s own mouth – in the midst of the ‘specious reasoning’ – that he explicitly presents the substantive view that acting in one’s own self-interest is never unreasonable.

Hobbes accedes to the Foole’s usage to engage him on his own terms. Since the Foole uses ‘reason’ to refer to the perspective of self-interest, a simple appeal to obligation in reply to him would be pointless: a change of subject, even. After all, the Foole does not care about what he ought to do morally: he cares about what is in his self-interest. While Hobbes could have just told the Foole that he ought to care about what he ought to do morally, this would have been unlikely to persuade the Foole. Perhaps Hobbes, like many moral philosophers, thinks (in my view mistakenly) that he needs something to say that will ‘win round’ an egoist such as the Foole by his own lights. In any case, it would still be pragmatically important to Hobbes to show that covenant-breaking is rarely in one’s self-interest, since he undoubtedly thought that, as a descriptive matter, people are strongly moved by their own self-interest.⁴¹ One of *Leviathan’s* key purposes was to convince people that an ordered, moral society is in their interests. It would be bad rhetorical strategy to refuse to engage the Foole on this point.

On this interpretation, the reply to the Foole is a more or less pragmatic attempt to convince egoists that it is generally in their self-interest to keep covenants. Such an attempt can still be largely successful even if we can imagine occasional cases where breaking covenant is in an agent’s self-interest.⁴² By contrast, on the rule-egoist interpretation, the reply to the Foole takes on the huge burden of showing that it could *never conceivably* be rational to break covenant, even from a solely egoistic point of view. For the rule-egoist, the binding quality of the laws of nature hangs on executing this task; the whole of Hobbes’ moral theory comes crashing down if this cannot be done. And ingenious as rule-egoist interpretations have been, the task is a very hard one to execute.

⁴¹ As such, Hobbes thought that to ensure actual compliance, it wouldn’t be enough to give a moral argument: a sovereign is needed to ensure that covenant-keeping lines up with self-interest, as Gauthier (1969: 88-9) notes. But contra Gauthier, this isn’t to deny that the moral argument is enough to establish genuine obligation.

⁴² The idea that the reply to the Foole may be more like a general case that covenant-keeping is in one’s interest than a hard and fast proof that the two could never diverge is not new; it has been urged by Kinch Hoekstra (1997). However, as noted above, Hoekstra thinks Hobbes agrees with the Foole that reason compels one to seek one’s own self-interest in cases where the two diverge; here, we obviously disagree.

I do not deny that there are traces of the rule-egoist reasoning in Hobbes' attempt to show in the reply to the Foole that adherence to the laws of nature tends to promote the agent's self-interest. In fact, I can largely agree with the rule-egoist when it comes to interpreting Hobbes' reasoning as to why compliance with one's obligations is (generally) in one's self-interest. What we disagree about is the role that this reasoning plays in Hobbes' broader moral theory. I have argued that in key passages Hobbes very clearly endorses the possibility of obligations that cut against the agent's self-interest – something which, it bears stressing, rule-egoism denies just as much as standard egoism does. So I do not believe that it is Hobbes' considered view that the genuine ground of our obligation is to be found in the egoistic considerations he appeals to in his reply to the Foole. Rather, the ground of our obligation is our covenanting.

3. The coherence of Hobbes' view

I think that my interpretation is consistent with the bulk of Hobbes' text, and has no obvious incoherence. This is enough to give it an advantage over alternative explanations. Of course, it may be open to substantive objections; that is not my topic here. Briefly, however, I want to respond to one argument that charges the view with genuine incoherence. It might be thought that if we have only self-interested reasons to covenant, and the point of covenanting is only self-interested, then we cannot have reasons to keep covenant when it is against our self-interest. For the keeping of covenant is only valuable to us as a means to an end of furthering self-interest.

I do not think that this follows. The view I am attributing to Hobbes is not that all reasons are egoistic, but rather that though there are non-egoistic reasons, such reasons apply to us only because we have made covenants, prior to which we have only egoistic reasons (and which, therefore, our only reasons for making which are egoistic). One can have reasons of type A to perform an act which gives rise to reasons of type B, without the type B reasons being themselves reducible to reasons of type A (though their origins can be explained by the type A reasons). Furthermore, the type-B reasons are not just outweighed by just any further contravening type-A reason. What it would take to outweigh the type-B reason is for the type-A reason to give oneself type-B reasons itself to be outweighed. But in the Hobbesian case, this would have to amount to us being better off outside the commonwealth, and for Hobbes this is never the case. It is our ongoing self-interested need to be part of the commonwealth that gives the non-egoistic obligations that arise from our membership a kind of sticking power.⁴³

To make this plausible, let me appeal to an analogy: playing sport. Imagine a person who plays sport for self-interested reasons: he wants the glory and public accolades that come with winning. These are his only reasons for engaging in sport. Nevertheless, such a person might think that, having chosen to play sport, he has undertaken an obligation not to cheat. His obligation not to cheat is not just based on the possibility that he will be found out: he thinks he would be obligated even if his cheating was guaranteed to be covert. He just thinks that by undertaking to participate in the game, he has taken on an obligation to play fairly. But in recognizing this obligation, he in no

⁴³ Thanks to Tom Feeney for pushing me to clarify this.

way sees his reasons for playing sport as being a general desire to further the good of everyone that plays.

Now, you may disagree on substantive grounds with this conception of reasons for playing sport. You may be an egoist about sport, who thinks that if cheating can advance your own self-interest, it is worth doing. At the other extreme, you may, like Onora O'Neill (1989:102-3), be a Kantian about sport, and hold that it is impermissible to play if your over-riding motive is to win, because your maxim is not universalizable. But maybe you are a Hobbesian about sport: maybe you think that your reason for engaging in sport is self-interested, but that in agreeing to play you take on obligations that might cut against your self-interest. And even if you are not, the point is just that the position involves no obvious inconsistency or irrationality.

If that's right, I suggest, then there is also nothing obviously inconsistent about the analogous position about social covenant. If we take seriously the possibility of *assuming* an obligation, there is nothing inconsistent about having non-self-interested reasons once one has covenanted, but having only self-interest reasons to covenant. The interpretation thus avoids conceptual incoherence. Whether Hobbes' position is mistaken will have to be settled at the level of substantive moral philosophy.

References

Abbreviations for primary texts

DCv Hobbes, *De Cive (On the Citizen)*. Translations are from the Tuck & Silverthorne edition (Cambridge: Cambridge University Press, 1998); paginations refer to the chapter and paragraph marked in that edition.

L Hobbes, *Leviathan*. Paginations refer to the chapter and paragraph marked in the Curley edition (Indianapolis: Hackett, 1994).

MS Kant, *Die Metaphysik der Sitten (The Metaphysics of Morals)*. Paginations are to the *Akademie* edition.

Other works cited

Darwall, S. (2000). 'Normativity and Projection in Hobbes's *Leviathan*,' *Philosophical Review*, 109/3: 313-347.

Gauthier, D. (1969). *The Logic of Leviathan*. Oxford: Oxford University Press.

----- (1979). 'Thomas Hobbes: Moral Theorist,' *Journal of Philosophy*, 76/10: 547-559.

Gert, B. (1967). 'Hobbes and Psychological Egoism,' *Journal of the History of Ideas*, 28/4: 503-520.

Hoekstra, K. (1997). 'Hobbes and the Foole,' *Political Theory*, 25/5: 620-654.

Kagan, S. (2000). 'Evaluative Focal Points,' in Hooker, Mason & Miller (eds.), *Morality, Rules and Consequences: A Critical Reader*. Edinburgh: Edinburgh University Press.

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

----- (1995). 'The Rationality of Rule-Following: Hobbes's Dispute with the Foole,' *Law and Philosophy*, 14: 5-34.

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

Ludwig, B. (2004). 'Thomas Hobbes – Recht, Unrecht und die Selbstverpflichtung der Menschen', in *Klassiker der Philosophie heute*. Stuttgart: Reclam.

Moore, S. (1971). 'Hobbes on Obligation: Moral and Political: Part One: Moral Obligation,' *Journal of the History of Philosophy*, 9/1: 43-62.

Nagel, T. (1959). 'Hobbes's Concept of Obligation,' *Philosophical Review*, 68/1.

----- (1970). *The Possibility of Altruism*. Princeton: Princeton University Press.

O'Neill, O. (1989). *Constructions of Reason: Explorations of Kant's Practical Philosophy*. Cambridge: Cambridge University Press.

Rawls, J. (1955). 'Two Concepts of Rules,' *Philosophical Review*, 64: 3-32.