

Hobbesian Public Reason*

Thomas Hobbes observed that human interaction naturally generates destructive conflict. To reap the benefits of human interaction while constraining the destructiveness of such conflict, many have urged that we need a conception of public reason, where "public reason" is understood as reasoning in a special way about a distinctively public set of issues. Many philosophers have remarked that the way we reason about distinctively public issues is (or, at least, ought to be) importantly different from the way in which a lone individual would (or ought to) reason about an issue that is not so distinctively public. For example, John Rawls argues that when reasoning about what he calls "constitutional essentials" citizens should not appeal to their full comprehensive views, but instead should deliberate only in terms of public values.¹ One might hope that a theory of public reason could explain this distinction between public and nonpublic reason, and, perhaps, explain it in a way that demonstrates both how public reason might be justified and gives us the resources to criticize any particular instance of such reasoning. A particularly Hobbesian motivation for such a theory is the idea that everyone's relying only on the individualistic mode of reasoning while interacting with one another guarantees destructive conflict, and a life that is "solitary, poor, nasty, brutish, and short."²

Hobbes notoriously argues that the best solution to this problem is to set up a sovereign whose authority is completely unconstrained. As Locke famously argued, such completely unconstrained authority runs a serious risk of creating a monster. David Gauthier has recently tried to remedy this problem, while taking very seriously the conflict Hobbes wanted to ameliorate. Gauthier has long been impressed with Hobbes's moral and political philosophy, so it should come as no

surprise that his account is deeply Hobbesian. Indeed, he only makes two modifications to Hobbes's account, although the modifications are quite significant.³ To reduce the obvious dangers of unconstrained authority, Gauthier argues that we should limit (a) the scope and (b) the content of the sovereign's authority. His proposed modifications are natural and tempting ones for anyone wanting to defend a Hobbesian theory of public reason, so his view provides a useful exemplar of the sort of approach I want to consider.

Gauthier's approach has much to recommend it as an improvement upon Hobbes's account. However, there is an alternative strategy for solving Hobbes's problem that Gauthier does not consider. Rather than focusing on the authorization of some public person, or sovereign, as both Hobbes and Gauthier do, this alternative strategy focuses on the authorization of some public set of principles. Whereas for Hobbes and Gauthier, the primary question for a theory of justice is "*Who* has the authority to settle these disputes?" the primary question for the proposed alternative is "*What* principles have the authority to settle these disputes?" This alternative account does not deny that we will very often need to appoint judges to settle our disputes, but urges that we need some shared background conception of public reason -- some principles of public reason -- to constrain the appropriateness of the arbitration of such judges. I argue that anyone motivated to move beyond Hobbes's imperialistic sovereign to Gauthier's constrained sovereign, should also be motivated to move beyond Gauthier's constrained sovereign to my proposed alternative, which gives primacy to public principles rather than to some sovereign, or public person. I also argue that this alternative theory will yield a theory (better: theories) that is (are) more plausible

than Gauthier's. In sections I and II, I present this alternative as a natural development in light of the accounts of Hobbes and Gauthier. Hobbes has an interesting argument, the regress argument, against my proposed alternative. However, anyone who restricts the sovereign's authority in the way Gauthier does will not consistently be able to deploy this argument; I argue for this in section III. In section IV, I contend that any Hobbesian theory of public reason faces a serious dilemma, and suggest that my proposed alternative stands a better chance of evading its force than either Hobbes's theory or Gauthier's theory does. Finally, I conclude with some brief remarks about what we should do if, at the end of the day, the case for my proposed alternative falls short, and *neither* it nor any Hobbesian theory of public reason can evade the force of this dilemma.

It is worth being clear at the outset that the most distinctive feature of Hobbesian accounts is their attempt to derive a conception of public reason from a conception of natural reason. This sets such accounts apart from at least two other prominent kinds of theories. Broadly Wittgensteinian views deny the Hobbesian assumption that there is any distinction between public reason and natural reason, maintaining instead that all reason is public reason. I briefly discuss such Wittgensteinian accounts, and Gauthier's response to them, in section I. One might, however, allow the Hobbesian assumption that there is an important distinction between public reason and natural reason and still reject the Hobbesian ambition to derive the former from the latter. This second sort of alternative to the Hobbesian account allows that public reason and natural reason are importantly distinct but hold that each is primitive, in the sense that neither one is derivable from the other.⁴ Rawls's recent work makes it clear that he falls into

this camp.⁵ Hobbesians reject the Wittgensteinian assumption that all reason is public and the Rawlsian assumption that it is misguided to try to derive a conception of public reason from a conception of natural reason. For present purposes I share the Hobbesian ambition of deriving a conception of public reason from a conception of natural reason, and in this way my proposal differs importantly from both Wittgensteinian and Rawlsian approaches. The aim of the present paper is to present an account that is better able to fulfill this Hobbesian ambition than either Hobbes's view or Gauthier's attractively modified Hobbesianism. More broadly, the aim is to try to develop the most plausible Hobbesian account of public reason without, at this stage, prejudging whether we should, in the end, favor the Hobbesian approach over a more Wittgensteinian or Rawlsian approach.

I. Hobbesian public reason without Hobbes's imperialism.

Hobbes famously argues that to avoid the conflict of the state of nature 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."⁶ We all must, in effect, supplant our private reason and judgment with the reason of the arbitrator, which thereby becomes public reason for us. Hobbes understands this arbitrator to be imperialistic; his judgments on any topic whatsoever must be taken as final. Gauthier argues that Hobbes's account provides "most of the materials needed for an account of public reason,"⁷ and that it is only in leaving the arbitrator unconstrained that Hobbes's account has gone seriously awry. Gauthier suggests that those in the state of nature would have good grounds for limiting the scope of public reason and that, in recognizing them, we might salvage Hobbes's account

simply by limiting the scope of the arbitrator's authority accordingly. In this section, my primary aims are exegetical. I discuss what Gauthier plausibly takes to be the key features of Hobbes's account of public reason, why he thinks the account needs to be modified, and his proposed modification. As a preliminary to the discussion of these elements of Gauthier's view, consider how Gauthier defends his view against those who argue that reason is essentially public.

An important presupposition of Gauthier's view is that there is a distinction between individual reason and public reason; otherwise we could not construct an account of public reason out of the materials of individual reason as Gauthier attempts to do. He recognizes, however, that this presupposition will be challenged by those who are led by certain Wittgensteinian considerations to the conclusion that all reasoning is essentially public. In light of this potential threat to his view, Gauthier briefly provides his own account of reason, and explains why such Wittgensteinian considerations do not really tell against his view. His response is worth reviewing, particularly since I suspect his response is not completely convincing, and since we may want to return to the Wittgensteinian account if no Hobbesian theory, including my proposed alternative, is able to meet the dilemma I press against such theories in section IV.

Gauthier argues that the core of reason is the capacity to act for reasons, a capacity that humans, some nonhuman animals, and many artificial bodies share. Following Dretske⁸ he emphasizes that reasons are "representations of some part of the world, as it is or as it might be."⁹ When an agent acts for reasons, we explain the agent's behavior in terms of the content of some such representations. We can then distinguish the capacity to act for reasons, which is a motivational

capacity that non-linguistic entities might even have, from rationality, which he suggests is a normative capacity that requires language. Rationality is the capacity to assess one's representations as reasons and be moved only by those that one finds appropriate.

On Gauthier's account, there is nothing incoherent about the supposition of an agent exercising rationality without a community. This contrasts sharply with the view that rationality involves language and language, as Wittgenstein showed, is necessarily a social practice. In response to the Wittgensteinian objection, Gauthier argues that while human beings are such that unless we are exposed to linguistic behavior in childhood we won't actualize our linguistic capacity, there is no reason, "to deny in principle that a teachable language could be developed by a solitary being with a linguistic capacity different in some respects from that which we actually possess."¹⁰ In this way he admits that rationality depends on language but denies that this justifies a communitarian conception of rationality by resisting the claim that language is essentially social. Gauthier admits this argument is not sufficiently developed, but for present purposes, we might grant him this point, at least provisionally.

Gauthier distinguishes this Wittgensteinian argument for an essentially social conception of rationality from what he takes to be an importantly different argument for that same conclusion. His discussion of what is both right and wrong about this second argument serves as a useful segue into his account of what Hobbes takes as a key source of the conflict that public reason is meant to solve. He offers Harold Brown's version of this second argument. On Gauthier's interpretation, Brown argues from the reliability of rationality and the fallibility of

individual agents to the conclusion that rationality is essentially social.¹¹ What Gauthier thinks is right about the argument is its recognition that an agent who fails to consider whether competent judges in her community share her assessment of the rationality of her actions is less than fully rational. Brown's mistake, in Gauthier's eyes, is jumping from this to the conclusion that she is not exercising rationality at all. Since, after all, the reason of an entire community is fallible as well, the difference between a lone reasoner and a community of reasoners is one of degree and not kind. Hence, Gauthier urges, we should recognize that rationality may be exercised asocially, though usually not as effectively as it is when exercised socially. Having considered and rejected both the Wittgensteinian argument and the sort of argument he attributes to Brown, Gauthier feels justified in assuming that rationality is not essentially public.¹²

Still, the fact that individual reasoners are especially fallible is important on the Hobbesian account. As Hobbes notes, "no one man's reason, nor the reason of any one number of men, makes the certainty."¹³ Given that individuals are fallible, they may differ in their judgments, and these differences may often lead to conflict. So Hobbes urges that the only way to avoid war is for each individual to supplant her own reasoning with the reasoning of some public person, insofar as the disputed matter goes. As Gauthier puts it, "the individual mode of deliberation, in which each person judges for herself what she has reason to do, is to be supplanted by a collective mode, in which one person judges what all have reason to do." (PR, 25) Hence, "rationality has a social dimension, not because of the unreliability or fallibility of individual judgment, but rather because of the controversies arising from that fallibility."¹⁴

Fallibility, however, is not actually the deepest problem that public reason is meant to solve. As Gauthier goes on to argue, Hobbes's view is actually that even if each individual were infallible, controversy would still arise and there could be in nature no single right reason to resolve it. In particular, Hobbes stresses that individuals, even the same individual at different times, will differ in their judgment of what is good and what is evil, and that this is a source of conflict. Importantly, on Hobbes's account, these differences need not arise because of the fallibility of reason:

Whatsoever is the object of any man's appetite or desire, that is it which he for his part calleth *good*: and the object of his hate and aversion, *evil*; and of his contempt, *vile* and *inconsiderable*. For these words of good, evil, and contemptible, are ever used with relation to the person that useth them: there being nothing simply and absolutely so; nor any common rule of good and evil, to be taken from the nature of the objects themselves, but from the person of the man, where there is no commonwealth...¹⁵

While some differences about goodness may stem from errors in reasoning, on Hobbes's account others may arise from differences in appetites. Finally, and most importantly, differences in the way that states of affairs bear on appetites can produce conflict, as "you and I may both find praise pleasant and criticism unpleasant; we shall therefore evaluate quite differently a state of affairs in which you are praised and I am criticized."¹⁶ Gauthier usefully notes that we need not accept the tight links between reason and individual good, and between individual good and appetite, to generate the problem that Hobbes had in mind. All we need is that, "the way in which two states of affair bear on a particular agent may

rationally affect her decision as to which to bring about, so that different agents might rationally reach different decisions."¹⁷ Gauthier argues that even if we have a considerably more sophisticated and less egoistic account of human psychology than Hobbes supposed, conflict could easily arise even among infallible agents. This conflict stems from a structural feature of interaction and could arise not only among various sorts of egoists, but among those with any of a variety of non-egoistic values.¹⁸

Having established that natural reason alone, the reason each of us has in the state of nature, cannot ensure that we will peacefully resolve our differences, Hobbes argues that the parties in that state should "set up for right reason, the reason of some arbitrator, or judge, to whose sentence they will both stand."¹⁹ Gauthier plausibly emphasizes that Hobbes does not simply have in mind an agreement based on each individual's judgment that he does best to accept the arbitrator's judgment, whatever it is, so long as others likewise agree; on this interpretation Hobbes's talk of setting up an arbitrator for right reason would be no more than "hyperbolic metaphor."²⁰ Rather, Gauthier suggests, Hobbes is making the deeper suggestion that in the context of the conflict we should actually transform the method of rational deliberation. Each individual's deliberation is supplanted by the deliberation of some arbitrator that they have authorized for this purpose, not by the individuals themselves. In authorizing the sovereign, each individual is understood "to give up her right to deliberate and will for herself."²¹ Importantly, as Gauthier points out, the sovereign is not simply to decide some particular controversy or range of controversies, but to judge all controversies among those who set him up and agree to accept his judgment. The sovereign

expresses right reason through the civil law, which is addressed to those who are already obliged to obey the sovereign. Civil law is understood as supplanting the individual judgments of citizens as to what is to be done. Numerous passages support this interpretation. For instance, Hobbes claims that,

The only way to erect such a common power...is, to confer all their power and strength upon one man, or upon one assembly of men, that may reduce all their wills, by plurality of voices, unto one will: which is as much as to say, to appoint one man, or assembly of men, to bear their person; and every one to own, and acknowledge himself to be author of whatsoever he that beareth their person, shall act, or cause to be acted, in those things which concern the common peace and safety, and therein to submit their wills, everyone to his will, and their judgments, to his judgment. This is more than consent, or concord; it is a real unity of them all, in one and the same person, made by covenant of every man with every man, in such manner, as if every man should say to every man, *I authorize and give up my right of governing myself, to this man, or to this assembly of men, on this condition, that thou give up thy right to him, and authorize all his actions in like manner.* This done, the multitude so united in one person...²²

Gauthier suggests that what creates a real problem for Hobbes, is that he also wants to identify public reason with equity. For example, Hobbes argues that,

Also if *a man be trusted to judge between man and man*, it is a precept of the law of nature that *he deal equally between them.* For without that, the

controversies of men cannot be determined but by war. He therefore that is partial in judgment, doth what in him lies to deter men from the use of judges, and arbitrators; and consequently, against the fundamental law of nature, is the cause of war.²³

This interpretation is further confirmed by the following passage:

...for in this consisteth equity; to which, as being a precept of the law of nature, a sovereign is as much subject, as any of the meanest of his people.²⁴

If we take the sovereign's reason to supplant our own, there would seem to be no room for us to apply a further standard of impartial dealing. On Gauthier's interpretation, Hobbes wants to counter the forces of chaos by having us, as citizens, see the sovereign's reason as supplanting our own. Given this picture, there would seem to be no room for Hobbes to recommend equity as a standard for the sovereign to meet; the only standard we have is whatever the sovereign decrees, and if his decrees are not equitable then, on Hobbes's own account, we must still regard those decrees as final. We no longer have any mode of reasoning from which we could criticize the sovereign's decrees; if we genuinely supplant our own reasoning with his, then we cannot help but regard his conclusions as our own.²⁵ The only apparent alternative would be to give the citizens a basis for "limiting the extent to which she accepts the subordination of her natural reason to the 'right reason' of the sovereign."²⁶ As Gauthier points out, a similar problem plagues Hobbes's suggestion that the civil laws are the sovereign's interpretation of the laws of nature, which again inappropriately constrains the sovereign. The sovereign's reason is made right reason by our authorization, regardless of whether

he deals with us equitably or whether his will interprets the law of nature. Here Hobbes is not taking seriously his own claim that the sovereign's "laws shall be unto us *whatsoever they be*, in the place of right reason."²⁷

Gauthier is a bit too quick here, though. There is still room for Hobbes to recommend equity as a standard that applies to the sovereign. That standard applies, first, in virtue of its not being in the sovereign's self-interest to disobey it, given the risk of a return to the state of nature via (irrational but perhaps inevitable) resistance to such inequity. It applies, second, in virtue of God's holding the sovereign to such a standard. For example, he calls God the "sovereign of sovereigns"²⁸ and there is no obvious reason to read this as mere metaphor. This is made quite clear when Hobbes notes that, "It is true, that sovereigns are all subject to the laws of nature because such laws be divine, and cannot by any man, or commonwealth, be abrogated."²⁹ While Hobbes's position may be rendered consistent in this way, Gauthier is still right to worry that the resulting view is unsatisfactory. The reliance on God is fairly obviously not going to provide a satisfactory foundation for a theory of public reason that can play an important role in contemporary, secular communities. Relying on the sovereign's self-interest in avoiding a return to the state of nature is not much more helpful. As Gauthier notes, the reliance on the sovereign's self-interest forces us to run an unacceptable risk of "creating a monster." Gauthier seems to be motivated by the same sort of worries that Locke rightly had about Hobbes's account; a sufficiently cruel and powerful sovereign might make the commonwealth worse than the state of nature.

Gauthier attempts to solve this problem by limiting the scope of public reason.³⁰ Hobbes's mistake, it would seem, was not in relying on the authorization of an arbitrator to eliminate conflict, but in giving that arbitrator unconstrained authority over everything. As Gauthier puts it, internalizing the public standard "must not issue in the complete subordination of the citizen's individual standard."³¹ Rather than give up their natural reason across the board, on Gauthier's account we should see each citizen as having good reason to agree "to authorize a public person to judge and will in her name, *on those matters and in those respects that significantly affect the interactions of the citizens and the public goods available to them.*"³² The idea is that many of the worries that motivated Hobbes to require the sovereign to deal equitably and interpret the laws of nature in formulating civil law stem from the sovereign's having such imperialistic authority. If we narrow the scope of his authority, Gauthier suggests, we should have less need to impose such standards.

In addition to this revision, Gauthier adds that there is "also an implicit limit, arising from the evident fact that however permanent we may wish the public person to be, and however much we may expect to benefit from authorizing him, our expectations may prove mistaken, and our wish unfounded."³³ It would be foolish and futile, on his account, to urge that citizens do not retain the right to reconsider their authorization, "should they find their wishes and expectations betrayed."³⁴ So long as they lack reason to reconsider, however, they are subject in both their deliberations and actions to the judgment and will of the public person.

Finally, Gauthier makes clear that he does not mean to imply that we should look for any actual, historical authorizing agreements. Instead, his idea is the familiar one of justifying particular institutions in terms of a hypothetical contract, arguing that the best justification for the acceptance of the judgment of some actual public person as supplanting our own, "would be that if one could have found oneself negotiating the terms of social interaction with one's fellows, in a state of nature from which public reason was absent, one would have acknowledged both the general need for such reason and the more specific authorization of the existing public person as the object of an agreement with one's fellows that it would not be unreasonable for each to make."³⁵ So it is in this context that we are to make sense of his Hobbesian account of authorizing an arbitrator.

II. Public reason without a "public person" - an alternative approach.

Before examining Gauthier's account of public reason, it is worth considering just what purposes an account of public reason is meant to serve. A striking fact about human interaction is that individuals disagree about what ought to be done with respect to matters that affect them all, which is a recipe for conflict. The Hobbesian gives us an account of the source of that conflict, in terms of human fallibility and a distinctively Hobbesian value relativism.³⁶ Whatever account one gives of the source of such conflict, though, it seems that the point of having an account of public reason is to answer the question, "how might we get along and rationally settle our disputes in spite of our differences?" Gauthier's suggestion is that the answer to this question is that we should adopt as our own standard of

reasoning, with regard to some specified range of distinctively public questions, the reasoning of some "public person."

This entails that the particular standard of reasoning citizens are to adopt as their public reason will be a direct function of who is authorized as a public person, as such a person will quite literally, it seems, "lay down the law" on how we ought to resolve such issues. Gauthier explicitly puts aside the question of what the particular characteristics of this standard will be, as that is "the province of the theory of justice"³⁷ as he wants to focus instead on the idea that it be internalized in the reasoning of the citizens. What is odd here is Gauthier's suggestion that just *what* the standard will be like falls within the province of the theory of justice; one would have thought, given Gauthier's account, that we would need no theory to determine that, at least in any direct way. All we need a theory to determine is *who* will count as a justified public person (which as we have seen, will be settled by whether she would have been rationally authorized by citizens in a hypothetical contract), and a theory of the scope of their authority (which Gauthier has provided in rough outline and which does not seem to be what he has in mind when he speaks of a "theory of justice"³⁸), and we will know that the standard of public reason is *whatever* that person decrees, so long as she stays within the specified scope of her authority. The theory of justice will fundamentally be answering a question about *who* gets the authority to establish the principles of public reason and only indirectly answering a question about just *what* standard should be adopted.

I have emphasized this point not primarily because I think Gauthier's suggestion about what falls within the province of the theory of justice is

misleading (though I think it is), but because thinking about his approach in this light naturally leads one to consider an alternative. As should be evident, the Hobbesian theory is first and foremost a theory of who should be the public person. It is only by endorsing someone as the public person, who in turn supplies us with a standard of reason on distinctively public issues that we generate public reason. A natural alternative to this approach would be to use the device of hypothetical agreement more directly to establish a public standard of reason. One might give a plausible account of what makes some particular conception of public reason justified that would look very much like Gauthier's account of what makes some public person count as justified. For example, one might hold that the best justification that one could be offered for the acceptance, of some particular, locally dominant, conception of public reason in the place of one's own individual reasoning, would be that if one could have found oneself negotiating the terms of social interaction with one's fellows, in a state of nature from which public reason was absent, one would have acknowledged both the general need for such reason and the more specific authorization of the existing conception as the object of an agreement with one's fellows that it would not be unreasonable for each to make.³⁹

Indeed, such a theory seems to have significant advantages when compared with the Hobbesian theory. First, it would deal more directly with the same considerations that Gauthier's theory inevitably considers only in a more roundabout way, as presumably the determination of whether some particular "public person" would have, under the appropriate hypothetical situations, garnered the support of the citizens will primarily be a function of the content of

the conception of public reason embodied in his judgments and will.⁴⁰ Second, since on this sort of account the citizens ultimately supplant their natural reason (on certain distinctively public issues) with the reasoning embodied in some publicly enshrined code⁴¹ rather than with the reasoning of some particular individual, the "risk of creating a monster" would be lessened as compared with Gauthier's approach. After all, a publicly enshrined conception of public reason is not suddenly going to transform itself in hideously tyrannical ways in the way some particular public person might. It is possible, of course, that a public code might come to be interpreted in ways that are unfair or even grotesque, but if this were to occur at least there would be some official code to which citizens could point in complaining; on Gauthier's account it is unclear how they could possibly complain, as the public person *defines* the standard for such issues. Third, Gauthier's account seems less psychologically feasible, as citizens would frequently have to reorient their conception of public reason with any changes in the perspective of the public person, not to mention that there will inevitably be transitional phases from one public person to the next.⁴² In light of these three considerations, I tentatively conclude that my proposed account is more plausible than Gauthier's.

Oddly, Gauthier does not discuss this alternative strategy for generating an account of public reason. Indeed, the only alternative approach he considers is one that appeals to "an independent ground of normativity." In particular, he argues that the only two ways to solve the problems generated by everyone following their natural reason is to find such an independent ground of normativity or "some inadequacy must be found in natural rationality, which may

be remedied...by the creation of a further ground of normativity." He then argues that the former strategy is closed to those who, like him, see the world as "normatively disenchanting"⁴³ and concludes that the latter strategy, of which his own is an instance, must be the proper solution. The problem is that his own account is not the only instance of the latter strategy; the account I have been sketching in this section seems to be another. So it would seem the reader is simply left to speculate as to what grounds Gauthier has for preferring his own account to some account more along the lines I have suggested. In the next section I will so speculate, and consider an argument that would be quite natural for the Hobbesian to use against the proposed alternative. I shall argue, however, that given the ways in which Gauthier's account differs from Hobbes's, he cannot consistently deploy such an argument.

III. The instability of a constrained Hobbesian Sovereign (or, "Why a Hobbesian can't domesticate the Leviathan").

Hobbes, unlike Gauthier, did consider the sort of proposal I have in mind. Yet the argument he offers against it is not available to those who wish to follow Gauthier in limiting the sovereign's authority. Consider how Hobbes replied to the suggestion that the commonwealth be ruled by a set of laws. Hobbes thought that human beings can never agree on one recognized interpretation of say, their constitution, so a judge will be needed for interpreting it for them. However, this is to make the judge the sovereign, and if we suggest that he also be subject to a still higher constitution, we would need yet another judge to interpret that constitution, and we would have generated an unacceptable regress.⁴⁴ Moreover, Hobbes takes his point to be quite general, as he argues that, "all laws, written and

unwritten, have need of interpretation...the written laws, if they be short, are easily misinterpreted, from the divers signification of a word, or two: if long, they be more obscure by the divers significations of many words..."⁴⁵ In virtue of these considerations, Hobbes concludes that we must settle upon some sovereign whose authority is completely unconstrained:

For to be subject to laws, is to be subject to the commonwealth, that is to the sovereign representative, that is to himself; which is not subjection, but freedom from the laws. Which error, because it setteth the laws above the sovereign, setteth also a judge above him, and a power to punish him; which is to make a new sovereign; and again for the same reason a third, to punish the second; and so continually without end, to the confusion, and dissolution of the commonwealth.⁴⁶

Hobbes's argument for absolute sovereignty as the only way to avoid the warfare of the state of nature relies essentially on this regress argument. His general point is that whenever there is a limited power in an orderly civil society, that power must be limited by a still greater power, and if that power is limited, then it in turn must be limited by a still greater power, and so on, ad infinitum, until we at last come to a power that limits all others but is itself absolute and unlimited.⁴⁷ He concludes that it is only when we have an absolute sovereign that we can have true government and escape the state of nature.⁴⁸

This suggests a tempting route for Gauthier to take in responding to the suggestion made in the previous section. Following Hobbes, Gauthier might argue that any publicly enshrined conception of public reason, like any law or rule, will need interpretation, so we will need some judge or arbitrator to

determine how the conception is to be understood. If we then try to limit the power of that arbitrator with some further set of rules, those rules will also need interpretation, so we will be set on a vicious regress. Any attempt to establish public reason through some official set of rules or ideals, if it is to be practically efficacious, must be converted into an establishment of public reason through an arbitrator. An attempt to instantiate the sort of theory sketched in the previous section will collapse into one that is essentially equivalent to Gauthier's.⁴⁹

At this stage, it is important to remember just how Gauthier's account diverges from Hobbes's. First, Gauthier, unlike Hobbes, limits the scope of the public person's authority; indeed Gauthier emphasizes this point as what makes his account a distinctive improvement upon Hobbes's. Second, Gauthier also allows that there is not only a limit on the scope of the arbitrator's authority, but also some tacit constraints on the substance of his decrees. As Gauthier puts it, "it would be foolish to urge, and futile to expect, that individuals not retain the right to reconsider their authorization, should they find their wishes and expectations betrayed." Otherwise we risk "creating a monster."

The trouble, as may by now be apparent, is that if the regress argument is effective against the alternative theory proposed in the preceding section, then it would also be quite potent against Gauthier's own account. Consider the limitation on the scope of the arbitrator's power. We might suppose that this limitation is even supposed to be made explicit. However that scope limitation is spelled out, it is going to require interpretation, insofar as we accept one of the main premises of the regress argument as Gauthier would deploy it against the account of the preceding suggestion.⁵⁰ Now we can either let the arbitrator

himself interpret the scope limitation however he wishes, in which case it is really no limitation at all, or we will need some new authority to interpret it, in which case we are off on Hobbes's regress.

Now consider the more substantive constraint, that the citizens may reconsider their authorization if they "find their wishes and expectations betrayed."⁵¹ The motivation here is that to sufficiently reduce the risk of "creating a monster" we must not only limit the scope of the arbitrator's power, but must limit its content within that scope. Just what is to count as a betrayal of the wishes and expectations of the body politic, or of some particular citizen? Even more clearly than in the case of the scope limitation, such a constraint will have to be interpreted, and seems even more capable of indefinitely many interpretations; at any rate, anyone who found Hobbes's regress argument to be generally plausible would be *very* tempted to deploy it here. Here Gauthier is explicitly leaving the judgment of whether such a betrayal has occurred in the hands of the citizens, explicitly giving them authority over the arbitrator.⁵² So it is hard to see how he could avoid the charge of being set on a regress, as long as he accepts the main premises of the regress argument as deployed against the account provided in the previous section. It would seem that his positive proposal entails that the regress argument is, in fact unsound, and hence no threat against his view or my own. Gauthier's account makes public reason less imperialistic and in that way more plausible than Hobbes's, but at the price of robbing the account of one of its chief defenses against the sort of alternative I am proposing. More generally, it is not obvious how any substantive constraint on either the content or the scope of the

public person's authority could consistently be adopted by anyone who was also deploys the regress argument.

Of course, this would only amount to an ad hominem attack on Gauthier's position vis-à-vis mine, unless some independent reason could be found for supposing that the regress argument is, in fact, unsound. The mere fact that Gauthier might be forced to assume that it is unsound to defend his own proposal might simply lead one to conclude "so much the worse for his proposal."

Fortunately, there are independent reasons to suppose that the regress argument is unsound. In particular, the argument's main premise, that all rules are inevitably subject to competing interpretations is, I think, demonstrably false. This can perhaps be best illustrated by considering how the argument, *if* sound, would vitiate Hobbes's own proposed solution.

As an anonymous editor for *Ethics* usefully pointed out, there is a real worry about how Hobbes could have solved his own problem.⁵³ The trouble is that if rules and edicts are incapable of constraining their interpretations, then the phrasings of the application or interpretation of the rule and edicts will presumably do no better. Suppose that the original edict is X, and the interpretation takes the form, "In circumstances C, rule X requires conduct A." Since this interpretation of the original edict is not on the Hobbesian view self-interpreting, it would seem that we have gotten nowhere by having a sovereign with absolute authority around to issue and interpret edicts. Assuming the regress argument is sound, Hobbes's own account falls prey to it.

This is a genuine problem for Hobbes. On the one hand, he must assume that the sovereign is capable of communicating with his subjects; otherwise it

would be nonsense to suppose that they could follow his commands and that he could do anything to ameliorate the conflict of the state of nature. However, in assuming that the sovereign is genuinely capable of communicating with his subjects, he seems to be supposing that some edicts or commands (or, at least, some of his clarifications/interpretations of those edicts or commands) are *not* particularly subject to competing interpretations - subjects, no matter how afraid of punishment, cannot obey commands that they cannot understand. However, if, as Hobbes seems forced to allow, some rules tend to elicit a similar interpretation from most people who are asked to interpret it, then it is no longer obvious that we need a sovereign. For the central tendency of interpretation of the majority and their predictable punishing responses to atypical interpretations might have the same force as the sovereign's threat of punishment; in each case, those initially inclined toward atypical interpretations will be deterred by the threat of sanctions, whether from an absolute sovereign or from "the people." Of course, in the state of nature, nobody will be rationally motivated to punish someone simply because they violated some rule; relying on their natural reason they will only be rationally motivated to punish someone when doing so would advance their conception of the good, and granting Hobbes his value relativism, those may be quite idiosyncratic. However, by transforming their reasoning in such a way that they come to *identify* with some commonly accepted and mutually advantageous set of principles, as suggested by my proposal, individuals might come to be rationally motivated to punish rule-breakers. There is, of course, the worry that this transformation itself will not be rational, which is parallel to the worry in Hobbes, as Gauthier interprets him, that it will not be rational to transform one's reasoning

to identify with the reasoning of the sovereign. I will return to this important worry below.

So the regress argument is unsound for the platitudinous reason that people are capable of communicating with one another. Unless people are being deliberately difficult or disingenuous, they more often than not know perfectly well what is being demanded of them when someone else commands them to do something in a language they speak. Moreover, even Hobbes would have to allow that this is true, and would be true even in the state of nature. Otherwise the sovereign could do nothing to ameliorate that state for the simple reason that his commands would be subject to endless misinterpretation, despite his best efforts to clarify his commands. This is an important result; if the regress argument were sound, it would tell against my own proposed account as well as Gauthier's.

IV. The Hobbesian Dilemma and the primacy of principle (or, "Why Hobbesian public reason puts the cart before the horse").

So far we have considered the Hobbesian account of public reason, which makes the authorization of a public person primary and derives standards of public reason from the decrees of that public person, and contrasted it with accounts which make the standards themselves primary. Some brief remarks have been made in favor of the latter strategy for generating theories of public reason, and we have seen how Gauthier would be unable to use the sort of strategy Hobbes might have used (the regress argument) to defend his approach versus the one identified in section two. Moreover, we have seen how the regress argument is unsound in any case, and hence no threat to my proposed account. Of course, Gauthier could always respond by denying that he would have mobilized the

regress argument in his defense.⁵⁴ While I suspect that, robbed of this strategy, Gauthier will not in fact be able to give anything approaching a compelling defense of his account's superiority over the alternative proposed in section two, in the present section I take this possibility seriously, and lodge an objection that cannot be dodged in this way. Moreover, the objection raised in the present section is one that *any* Hobbesian theory of public reason will have to take seriously.⁵⁵ Furthermore, the objection is connected in interesting ways to the explanation of why Gauthier could not consistently use the regress argument.⁵⁶

The Hobbesian begins by arguing that certain features of the state of nature would make conflict inevitable. As long as everyone continues to rely on their natural reason, this conflict will continue; natural reason is simply incapable on its own of resolving most of these conflicts. The Hobbesian suggests that we allow our natural reason to be supplanted (at least on some questions) by the reason of some public person who we authorize as such. The Hobbesian faces a dilemma. Either the public person's authority is limited or it is not. If it is not, then, in Gauthier's terms, we run an unacceptable risk of "creating a monster"; the risks of giving one person (or body), who is, after all, still simply operating with her own natural reason however much we pretend otherwise, seem to be too great, so that authorizing such a public person would not be rational.⁵⁷ If it would not be rational to authorize such a public person, then the state of nature story does not do the work it is supposed to do; it does not provide a rational reconstruction of how the public person might have come about that serves to legitimize her.

If, on the other hand, the public person's authority is limited, then we run into another problem. If such a limit is to be effective in reducing the risk that we

might "create a monster," it would must either be limited by some other authority (who in turn would need to be somehow limited, insofar as the first horn of the dilemma is potent; note the connection between this horn of the dilemma and the regress argument) or by the citizens.⁵⁸ If it is limited by the citizens, then they presumably must rely upon their private, natural reason to impose such limits, since it is hard to see how they could rely on the sovereign's reason to limit the sovereign. If, however, the citizens must rely on their natural reason to keep the public person in check, then we risk being thrust back into the state of nature, given the inability of people relying simply on natural reason to sort out their differences peacefully. The Hobbesian would simply seem to have moved their disputes up a level from being disputes about what they ought to do to disputes about whether the sovereign's decree that they should do something (or reason in a certain way) is legitimate. Given the Hobbesian assumption that natural reason will, in general, be of little use in sorting out our differences it is hard to see how the conclusion that we would constantly risk being thrust back into the sort of conflict we found in the state of nature could successfully be resisted. Call this the Hobbesian Dilemma.

Keeping this dilemma in mind, it should not be surprising that Hobbes both wanted to give the sovereign absolute sovereignty and at the same time, hold him to some standard. Hobbes both wanted the sovereign's authority to be absolute, so that he might dodge the second horn of the dilemma, but he also wanted to require that the sovereign be equitable, so that he could dodge (or at least blunt the force of) the first horn of the dilemma. Gauthier limits both the scope and the content of the public person's authority, so he may avoid the first

horn of the dilemma, but it is unclear how he is supposed to avoid the second, given that he shares Hobbes's pessimism about the ability of natural reason to sort out differences.

There presumably are many moves that a Hobbesian might make to avoid or remove the force of this dilemma. Let us suppose, however, as I suspect, that whatever epicycles they might go through, Hobbesian views like those of Gauthier and Hobbes will not be able to disarm this dilemma; what conclusions should we then draw? One might suppose that the Hobbesian has set themselves an impossible problem, and go about articulating a more manageable problem for theories of public reason to solve.⁵⁹ This would just change the subject, and though this might be necessary at the end of the day, one might have hoped that some other strategy for solving the Hobbesian's problem would bear some fruit.⁶⁰

Not surprisingly, I want to press further the suggestion that the alternative offered in section two offers some real progress in this regard. It is important to stress that the contrast between these Hobbesian strategies for generating theories of public reason and the alternative I have offered is *not* that one simply involves principles and standards while the other involves authorizing an arbitrator or public person. Both accounts are going to involve both principles and standards, and the authorization of an arbitrator or public person to settle disputes. The difference is that Hobbes and Gauthier make the authorization of the arbitrator primary and the standards of public reason are derived from the arbitrator's will, while the alternative I am pressing makes the standards of public reason primary and derives when and how arbitrators should be authorized from those standards. Essentially, my diagnosis of inability of Hobbes or Gauthier to disarm the

Hobbesian Dilemma is that they gets things backwards; so long as the authorization of a public person is primary the dilemma will be inescapable.

Suppose we had developed a theory of public reason along the lines suggested in section two. Presumably, then, for a given society, we could use that theory to generate a set of principles to guide the members of that society in their deliberation on a certain range of distinctively public issues, as well as being able to use that theory to assess the principles tacitly governing their deliberation already. The idea would be that, ideally, on that range of distinctively public issues the members of the society would supplant their own natural reason with the reason embodied in the principles generated by our theory, just as the Hobbes and Gauthier would have them supplant their natural reason with the reason embodied in the dictates of the public person, at least where that person has been legitimately authorized. So far, no reference has been made to the authorization of an arbitrator, so at this stage both horns of the dilemma are simply irrelevant, as they both involve worries about the scope and content of the arbitrator's authority.

However, it is quite likely that even in an ideal society governed by such a conception of public reason, there will be cases in which even reasonable people sincerely following the public conception of public reason governing the society will reasonably disagree. In such cases, there will be good reason to select an arbitrator to settle the dispute, as everyone can agree that its being settled one way or the other is better than its not being settled. Presumably such an arbitrator's power will not be imperialistic in scope as Hobbes's sovereign was, nor will the arbitrator be able to decide the manner in whatever way she wants. So the first horn of the dilemma pressed against the Hobbes and Gauthier is easily avoided -

any arbitrators that are legitimately authorized will be appropriately contained so that the risk of creating a monster will not loom large. It would seem, however, that by limiting the arbitrator's power in this way that we will be impaled on the second horn of the dilemma just as Gauthier was, and for just the same reason.

It only seems this way, though. The crucial difference is that whereas when Gauthier's citizens assess whether the arbitrator is overstepping her bounds or betraying the trust of the citizens they must rely on their natural reason,⁶¹ the citizens in the ideal society we have been considering would instead rely upon public reason, as defined by the principles that define the society's conception of public reason. The reason the second horn of the dilemma is so threatening for Hobbes and Gauthier is that one of the primary reasons we needed to develop a conception of public reason was because natural reason was so ineffective at peacefully resolving disputes, so if we have to rely on natural reason again but just at a higher level in assessing the legitimacy of the public person, we seem to run a substantial risk of being thrust back into the state of nature. On the proposed account, however, this is not a problem, as the parties have an understanding of public reason that is distinct from "whatever the arbitrator wills" and they can rely on that understanding, rather than their natural reason, to assess the legitimacy of the arbitrator's decision-making. Insofar as the account of public reason is not as impotent as natural reason at resolving conflicts, this would not run the same risk of a return to the state of nature.⁶²

It might be helpful to take a more concrete, if somewhat forced, example. Apparently, the principle, "You do not give meat to a man who has a full pot" is an important one for members of the G/wi Bushmen of the Kalahari desert. The

principle seems to make a good deal of sense for the G/wi, as they have no means of preserving or storing meat, so that giving someone more than they need would be wasteful.⁶³ Let us assume for the sake of argument that such a principle would play an important role in an ideal conception of public reason for the G/wi, and that it does in fact play that role. That is, let us assume that the G/wi actually have supplanted their natural reason with a conception of public reason that is ideal for them, and that this principle is part of that conception.⁶⁴ A controversy might arise in a particular case as to whether someone gave meat to someone "with a full pot" since it might not be clear whether the person who was given the meat really needed the meat.⁶⁵ Someone else whose child is severely malnourished might therefore raise a complaint that this person instead should have given them the meat. The controversy might be peacefully resolved without an official arbitrator being established, depending on the case and how rich the prevailing conception of public reason is. We might suppose, for example, that the person who gave the meat to the person "with a full pot" does not deny that the person was in less need than the child in question. However, they might defend their gift by appealing to a (perhaps tacit) higher-order principle that justifies the principle in question, e.g. a principle that communal resources should be used and produced in a maximally efficient way, as well as by appealing to other relevant principles in the conception. They might argue that the reason they gave the extra meat to the person they did was because that person had shown great courage and strength in a recent skirmish with another tribe, and that the general practice of rewarding such courage and strength provides incentives for such attributes, which in turn help keep the community and its resources safe from attack. The first party may

or may not find this defense plausible. The important point is that the two partisans not only could discuss the case in light of the principles of public reason they accept, but they might also be able to resolve controversies about how those principles should be interpreted, or even revised, in terms of the conception itself, rather than being forced to fall back upon their natural reason. This is crucial because a natural response for Gauthier and Hobbes to make to my proposal is to argue that it faces a problem exactly parallel to the one I pressed against their views when it comes time to apply those principles. Insofar as we might have the resources from within the conception to assess any component of it in terms of its coherence with the rest, this objection is ill-founded.⁶⁶

A related worry is that we must forge the principles in question for ourselves, and we would then have to judge, without the benefit of public principles, the authority of those principles, and doing so would seem to present just the sort of problem I pressed against Gauthier and Hobbes. This is a more difficult case than the one discussed above, because there we had assumed that a conception had already been worked out and accepted and only needed to be interpreted or revised; hence we could appeal to public reason to assess public reason. In forging the principles in the first place, no analogous move seems available. Of course, talk about "forging" the principles from scratch is just a reference to a thought experiment; we are interested in a hypothetical agreement. This does not defuse the objection, though, as we must be able to tell a story about this hypothetical agreement that is plausible.

However, given the way that such a conception might evolve gradually over time, it is not clear that this is a fatal objection. A given community might,

over time and through a process of trial and error, come to settle on a set of practices that "work" reasonably well for them. Moreover, if acceptance of those practices was required for group membership, and its significant rewards, then it might be rational for each individual to accept those practices in the deep, Gauthier/Hobbes, sense of *transforming* their reason in the appropriate way. Such practices might embody a tacit conception of public reason, and that conception might even be made explicitly at some point, say if the community came to face some new problem and were unsure how to proceed. The key point is that principles as embodied in a community's practices can evolve over time with input from everyone in a way that a sovereign cannot; a community can quite literally pass on to succeeding generations those practices that are effective, but they can only pass on the sovereign's successor, who may or may not choose to heed the collective wisdom she inherits.

Gauthier might try to co-opt this account, urging that it is through just this sort of trial and error over the generations approach that we might imagine on his account explaining how we come to have standards for assessing when our "wishes and expectations are betrayed." This does not seem to be what Gauthier actually intended; he presents his view in a way that suggests that the arbitrator's deliberations themselves *constitute* public reason while we have an external standard (our "wishes and expectations") which is not in itself constitutive of public reason, by which we gauge the arbitrator's deliberations. Were Gauthier to revise his view and make the move being considered, he would have to concede that public reason is, at least in part, constituted by the community's practices and the principles tacit in those practices; otherwise citizens would have to fall back

on their ineffective natural reason in assessing their arbitrator. Once this concession has been made, however, it is unclear how we can give the role of the arbitrator the sort of primacy that was so distinctive of the views of Gauthier and Hobbes. If these practices can be constitutive of an effective conception of public reason with respect to arbitrator-evaluation, then why not with respect to a range of other issues? After all, many of the reasons we have for disapproving of a given arbitrator will be reasons to disapprove of certain practices. Once this much is conceded, it is hard to see how one would resist the conclusion that at least some human societies could get by without an arbitrator at all, or only with various arbitrators set up on a short-term basis in light of the prevailing conception of public reason to deal with short-term problems. At this point Gauthier's account would look less like Hobbes's account than he apparently intended, and the distinction between his view and mine would blur.

Of course, our parties might not be able to resolve their conflict on their own. They might then be led by public reason to select some arbitrator to resolve the issue, and also be led by public reason to abide by whatever the arbitrator decides, so long as certain conditions, specified by their shared conception of public reason, are met. For instance, if it could be shown that the arbitrator accepted a bribe, or considered information that would obviously be irrelevant according to the shared conception of public reason, either party could legitimately complain that the arbitrator's decision should be seen as void.⁶⁷ Citizens might first try to work out their differences on their own by direct appeal to the principles of public reason embodied in their community, sometimes even revising those principles in light of themselves, in a reflective equilibrium-style

methodology. When this strategy is ineffective or yields ambiguous results, they might still have the resources within the conception to settle the issue insofar as the conception could lead them to appoint a neutral arbitrator to consider the evidence carefully and decide the issue one way or the other, as it will be important for it to be settled. When they do select an arbitrator, their selection of one, and their assessment of her decision's legitimacy, could itself be guided by public reason, rather than natural reason. In this way we might avoid the second horn of the dilemma that impales accounts giving primacy to a public person.

Supposing that the more or less ideal principles of public reason that govern a society could be rather straightforwardly applied, Gauthier and Hobbes might complain that there will still be a problem with compliance. Many members of the society may not be inclined to comply with the dictates of public reason, either because they do not track their best (natural) judgments and so yield "wrong results" or because they cannot count on others not to view the results in this way. With regard to the first worry, it does not seem to take the point that individuals are supposed to have genuinely *transformed* their natural reason into public reason on some specified range of distinctively public issues. The idea is that they no longer think about those issues simply in terms of their asocial, natural reason; rather their judgment in such cases is always colored by their understanding of the dominant conception of public reason. It would seem that the issue of public reason's yielding a result that does not track their best judgments would simply not arise.⁶⁸ Furthermore, it would not seem fair for Gauthier and Hobbes, at any rate, to become deeply skeptical about the possibility of such transformations, given that a very similar transformation plays such an

important role in their own accounts. Why should the fact that the conception of public reason that is to supplant natural reason is drawn from, say, the background culture, rather than from the decrees of the sovereign, make the assumption of a transformation of individuals' reasoning any *more* implausible?

With regard to the second worry, it would essentially seem to be a worry that not everyone has genuinely abandoned their natural reason on the relevant public issues, at least insofar as the preceding reply to the first worry is plausible. This may well be a genuine problem in many real cases, but it is not obviously a fatal one. In particular, those who have not transformed their reason in the appropriate way might be relatively easy to spot and sanction if they flagrantly disregard the dominant conception of public reason. Further, in many cases where this is a serious problem, there may be even more subtle ways to screen individuals to be relatively sure that they have transformed their reason in the relevant way. Such screening might then reassure the remaining individuals enough that they would not abandon compliance. At any rate, it is hard to see why this would present any *more* of a problem for the proposed account than it would for a view like Gauthier's.⁶⁹

V. Conclusion.

The Hobbesian has identified a real problem worth taking seriously, but the proposals of Hobbes and Gauthier do not represent the most plausible solutions to this problem. As I argued in section two, we have reason to favor what might be called the "principle-first" approach that I am defending. In section three, we saw that Gauthier, unlike a more imperialistic, absolutist Hobbesian, would not be able consistently to use the regress argument to defend his view against this alternative.

Moreover, I argued in that section that the regress argument is unsound in any case. So far, this does not preclude his effectively arguing that his conception is more plausible than the one offered in section two on some other grounds, but it is not clear how such an argument would go. Nor is it clear how someone sympathetic to the Hobbesian approach would reply to the sorts of considerations brought forward in favor of the alternative proposed in section two. Of course, I have only given a very sketchy account of this alternative, but one might reasonably worry that disarmed of the regress argument that a more Hobbesian account of the sort Gauthier offers will not be able to supplant it.⁷⁰ Gauthier could abandon the modifications he made to Hobbes's account, and then consistently mobilize the regress argument, but to do this would reintroduce the worries he initially had about authorizing an imperialistic arbitrator.⁷¹ Finally, in section four, I argued that any Hobbesian theory of public reason faces a deep dilemma in virtue of its giving primacy to the authorization of some public person rather than to a set of principles. Further, I argued that the alternative proposed in section two is able to avoid this dilemma in virtue of its giving primacy to principle rather than to the authorization of a public person. Given that the dilemma facing the Hobbesian is quite deep and damning, the ability of the proposed alternative to avoid this dilemma would speak heavily in its favor.

The dilemma pressed is quite formidable, though, and it *may* turn out that both the proposed approach and the more Hobbesian approach exemplified by Gauthier are unable to evade its force completely. If so, then the dilemma may speak in favor of abandoning the Hobbesian ambition of deriving a conception of public reason from a conception of natural reason, and instead taking a more

Rawlsian view of public reason. Alternatively, we might reconsider Gauthier's assumption that not all reason is public reason, perhaps returning to a more Wittgensteinian account.⁷² In any event, we should be skeptical about attempts to generate a theory of public reason that draw their inspiration from Hobbes until we have an account of how they can evade the Hobbesian Dilemma.

*Many thanks to Fred D'Agostino, Robin Flaig, David Gauthier, Christopher Morris, Geoff Sayre-McCord, Gerald Postema, an anonymous referee, and the editors of *Ethics* for their many helpful comments and suggestions.

¹John Rawls, Political Liberalism (New York: Columbia University Press, 1993).

²Thomas Hobbes, Leviathan (New York: Collier Books, 1962), p. 100.

³It is worth noting that in the paper under consideration in the present essay, Gauthier seems to be pushing a view of public reason that is closer to Hobbes's own than he has in previous work. For example, in an earlier paper on constitutionalism, Gauthier notes that, "Law, for Hobbes, provides a *public reason*, authoritative over the private reason of each individual. Hobbes's account of the need for public reason is a version, albeit an extreme version, of the general argument I have given...we need accept neither his dismal vision, nor his insistence on a 'sovereign governor,' to adapt Hobbes's thought about law to my purpose." (David Gauthier, "Constituting Democracy," The Idea of Democracy, ed. David Copp, et. al. (Cambridge: Cambridge University Press, 1993), p. 326. In the paper that serves as the focal point of the present argument, Gauthier by contrast seems committed to Hobbes's dismal vision and his insistence on a sovereign governor, with the two qualifications to be discussed in the text. We need not interpret Gauthier as having modified his view or contradicted himself, though. Rather, we can see him as focusing on two different problems, one that is faced by humans in the more nasty state of nature that was Hobbes's nightmare, and, in the earlier essay, one that is faced after that problem has been solved in a Hobbesian manner but in which other kinds of problems threaten. This interpretation is suggested by Gauthier's remarks in the earlier paper, where he notes that, "I have contrasted two contexts of interaction; we should, however, consider the possibility of a passage from one to the

other." (Ibid., p. 317) Oversimplifying, the idea is that we begin deriving the a fairly modest conception of the reasonable from the rational and escape Hobbes's nightmare, and we then use that conception of the reasonable to bootstrap our way to a more robust notion of the reasonable. A crucial move in this bootstrapping is the development of what Gauthier calls "civic friendship" - see Ibid., pp. 318-319 for useful discussion. Indeed, one suspects that any plausible Hobbesian account of public reason that tries to cohere at all with the rather sophisticated forms of public reason that we find in contemporary societies will make similar "bootstrapping" moves.

⁴Of course, a third logical alternative would be to allow the distinction and maintain that natural reason is derivable from public reason, but not vice versa. I overlook this sort of view in the text since it seems so bizarre and I know of no views in the literature that correspond to it, but mention it here for the sake of completeness.

⁵Rawls puts this point in terms of "the reasonable" and "the rational" (where the former corresponds roughly to public reason and the latter corresponds roughly to natural reason), arguing that any attempt to derive the reasonable from the rational wrongly suggests that "the reasonable is not basic and needs a basis in a way the rational does not." (Rawls, p. 52)

⁶Hobbes, p. 42.

⁷David Gauthier, "Public Reason," Social Philosophy and Policy, (1995), 19-42 at p. 36.

⁸Fred Dretske, Explaining Behavior (MIT Press: Cambridge, 1988), p. 80.

⁹Gauthier, "Public Reason," p. 19.

¹⁰Ibid., p. 21.

¹¹See Harold Brown, Rationality (Routledge: London, 1988), p. 187.

¹²For present purposes I can for the sake of argument grant Gauthier's claims about Brown, but I should register in passing my deep reservations about those claims. Importantly, Gauthier's claim

that Brown's argument is "very different" (Gauthier, "Public Reason," p. 22) from the Wittgensteinian argument that Gauthier sets aside is quite dubious; Brown's argument is actually quite similar to the Wittgensteinian argument. The real force of Brown's argument is that for us to use *theoretical* (not practical) reason we must be able to make sense of the idea that we are applying rules, and that we can only make sense of this idea if we can make sense of the possibility of error in our application of those rules. Now if we thought that such rules existed transparently in Plato's heaven for all to see and check themselves against, this would be no problem, but this seems implausible (see Brown, p. 188). The only remaining plausible option, Brown urges, is for us to rely upon the assessments of other members of our community to determine whether we have appropriately applied a given rule; hence reasoning is essentially public. *Contra* Gauthier's suggestion, then, Brown's argument is quite similar to the Wittgensteinian argument he sets aside, and quite different from the characterization Gauthier gives us of Brown's argument, which is a rather simple-minded appeal to the unreliability of individual judgments, rather than an appeal to the much deeper suggestion that we could not even make sense of the possibility of erring in our application of a rule outside a community. So the argument Gauthier refutes is a scarecrow of his own construction.

It is also worth noting that Brown's argument is about theoretical reason and hence is about the rationality of beliefs not actions, though Gauthier simply assumes without argument that "we may surely extend his account to apply to actions." Once it is clear how essential the notion of correctly applying a rule is to Brown's argument, this is not so obvious; we would need to be convinced that there must be an appropriately analogous kind of rule-application involved in our practical reasoning.

¹³Hobbes, p. 42.

¹⁴Gauthier, "Public Reason," p. 24. Gauthier actually claims at some points that the fallible will of a natural person "comes, by our agreement, to be accepted as the infallible will of an artificial person." (Ibid., p. 34) He moves, in other words, from seeing the sovereign's reason as our own to seeing it as infallible, without comment. The main trouble with this move is that if we really come to see the sovereign's reason as infallible, we will find it very difficult to make sense of situations in which the sovereign changes his mind. If the sovereign really comes to hold something that contradicts what he once held, then we cannot intelligibly hold that he was both infallible then and is infallible now. Furthermore, there are passages in which Hobbes makes it quite clear that the sovereign is not really infallible (see, for example, Hobbes, p. 435). While this does not strictly speaking contradict Gauthier's claim that Hobbes wants us to see the sovereign's reason as infallible, it does seem odd that he would emphasize (or even mention) the sovereign's fallibility if he wanted to convince his readers that they ought really to regard their sovereign's reason as infallible. After all, it would be easier for his readers to genuinely believe that their sovereign was infallible if he did not remind them of his fallibility. Why suppose he would bring something to his readers' attention that would make it more difficult for them to regard their sovereign in this way? Furthermore, Gauthier does not need to make this claim to motivate his proposed alternative, as my discussion in the text should indicate. Thanks to an anonymous referee for bringing this point to my attention.

¹⁵Hobbes, pp. 48-49.

¹⁶Gauthier, "Public Reason," p. 28.

¹⁷Ibid., p. 29.

¹⁸This is illustrated by a recent paper by Jeremy Waldron, who provocatively argues that Kant's account of legal authority is aimed at a problem that is structurally isomorphic to Hobbes's. He argues that,

The premise of Kant's account is that, in the absence of legal authority, we must expect that individuals will disagree about right and justice and that this disagreement will lead to violent conflict...Kant's position is reminiscent of that of Thomas Hobbes: individuals fight in the state of nature, and the consequent war of all against all can only cease when people submit to a unitary sovereign. The propensity of individuals to make judgments of value does not mitigate Hobbesian conflict. On the contrary, value judgments are directly implicated in the struggle for survival.

Jeremy Waldron, "Kant's Legal Positivism," Harvard Law Review 109, 1996, p. 1545. As Waldron notes, this is a surprising view for Kant to take, given his characteristic theory of the "categorical simplicity of duty" (Waldron, p. 1547). The source of conflict in Kant's account, Waldron argues, is that there will be much controversy in the application of the concepts of possession Kant discusses (empirical and intelligible possession). Waldron argues that Kant's account can plausibly be supplemented with the observation that there are in fact massive disagreements about which theory of justice is the most plausible, nor do these differences show any sign of going away. For present purposes, the essential point is that a moral philosopher with a theory of value as far removed from Hobbes as Kant was worried about a very Hobbesian problem. The Hobbesian problem is *not* idiosyncratic to the Hobbesian theory of value, though other value theories may provide resources for solving the problem that are unavailable to the Hobbesian.

¹⁹Hobbes, p. 42. Throughout the text I follow Gauthier in using 'public person', 'arbitrator', and 'sovereign' interchangeably. A more fully developed theory might make finer distinctions between these, perhaps emphasizing the difference between legislative, judicial, and executive officials, but for present purposes we may abstract away from such worries.

²⁰Gauthier, "Public Reason," p. 25.

²¹Ibid., p. 31.

²²Hobbes, p. 132. Gauthier argues for this interpretation at length in his The Logic of Leviathan (Oxford: Clarendon Press, 1969). Of course, one might reasonably take issue with this as an interpretation of Hobbes. However, even if Gauthier's interpretation ends up being controversial Hobbes exegesis, the view he presents *as* Hobbes's is still an interesting position, worth considering in its own right. My primary interest in the present paper is with a family of views (the view of Gauthier's Hobbes, Gauthier's view, and my own) that are properly thought of as Hobbesian in virtue of their responding to Hobbes's problematic. Though, as should be obvious, I think much can be said in favor of Gauthier's Hobbes as sound exegesis, the main lines of my argument should remain of interest if I am wrong about that.

²³Hobbes, p. 120. The idea here is that if *any* man who judges the conflict of other men (and the sovereign is surely going to be doing this, either directly or indirectly as we see in such passages as, "But the sovereign is already agreed on for judge by them both, and is therefore either to hear the cause, and determine it himself, or appoint for judge such as they shall both agree on." Hobbes, p. 182) is required by reason to deal equitably with them, since inequity will lead to further conflict, risking a return to the violence of the state of nature.

Of course, Hobbes does not think we may disobey the sovereign simply because he behaves inequitably; such disobedience would also risk a return to the state of nature. However,

that we may not resist the sovereign because he has failed to live up to this standard does not mean that there is no sense in which he is required by reason to do so. The standard still applies to him, even if we may not enforce it upon him, as I discuss in the text below.

²⁴Hobbes, p. 254.

²⁵Note that this does *not* amount to regarding the sovereign's reasoning as infallible, as Gauthier sometimes suggests. For I may regard my own reasoning as fallible, but still recognize that I cannot intelligibly step outside of my reasoning altogether and rationally reject its conclusions; for any such rejection to be rational, I would have to be relying upon my own reasoning, in some way, as its basis. We cannot coherently disagree with the sovereign, once we identify our reasoning with his.

Moreover, this account can explain why we cannot rationally disobey the sovereign (so long as he does not sentence us to death or demand us to incriminate ourselves). Gauthier brings this point out nicely in terms of Kurt Baier's distinction between rationality's reflective and executive tasks. The reflective task of reason is to assess the fit between one's representations of the world as it is and might be, and one's actions; the executive task is to be guided by that assessment. Gauthier mobilizes this distinction to explain that Hobbes is claiming that, "an agent who has agreed to accept the judgment of some individual or collective body as if it were right reason, and who then disregards that judgment in deciding what to do, is failing in the executive task." (Gauthier, "Public Reason," p. 26) Hence we can explain why, once a subject has identified her reasoning with that of the sovereign, it will be impossible (so long as this identification is maintained) for her to disagree with the sovereign and irrational to disobey him. So Gauthier can mobilize support for his account in the way described in the text, without

claiming that Hobbes wants us to regard the sovereign's reasoning as infallible, which would generate the difficulties discussed above (in footnote 12).

²⁶Ibid., p. 35.

²⁷Thomas Hobbes, English Works, ed. Sir Wm. Molesworth, (London: John Bohn, 1841), p. 194.

It is no coincidence that Hobbes continually makes such a strong claim, nor could he easily abandon it and keep his overall account intact. He argues that any attempt to apply some standard to the sovereign's behavior would tacitly put that standard above the sovereign, and hence he would not be genuinely sovereign, and thus could not solve the problem for which he was authorized. This element of Hobbes's account will become crucial below. Jean Hampton usefully labels this the "regress argument." See her Hobbes and the Social Contract Tradition, (Cambridge: Cambridge University Press, 1986), p. 98.

²⁸Hobbes, p. 251

²⁹Hobbes, p. 240. Thanks to an anonymous referee for prompting me to realize the importance of this element of Hobbes's view.

³⁰There are echoes here of a comment made by Gauthier in a previous paper: "The construction of a plausible contractarian theory of law will then depend on developing an account of the scope and standard of public reason. I shall not, of course, attempt to offer such an account here."

David Gauthier, "Thomas Hobbes and the Contractarian Theory of Law," Canadian Journal of Philosophy supplementary volume 16, 1990, p. 25.

³¹Gauthier, "Public Reason," p. 36.

³²Ibid., p. 37. Gauthier allows that 'significantly' is "deplorably vague" but lets this stand as an initial approximation.

³³Ibid., p. 37.

³⁴Ibid., p. 37. Gauthier also puts the point here by reminding us of Hobbes's claim that there is no right reason constituted by nature, and hence "to take the reason of one individual or assembly as if it were right reason is to risk creating a monster."

³⁵Ibid., p. 38.

³⁶Following Rawls, one might instead (or additionally) explain much of the disagreement found in modern industrial democracies in terms of the rather divergent comprehensive theories that color the reasoning of individual citizens different ways (which is consistent with most citizens not being firmly in the grip of any one particular comprehensive view, but rather, for example, endorsing bits and pieces of several different ones and trying to construct their own comprehensive view from those bits and pieces, or torn between several such theories and trying to settle on one). One need not endorse the sort of Hobbesian value relativism that Gauthier seems to endorse to adopt this view. The crucial point is that reasonable citizens may differ on these matters, and do so reasonably; one might see this as an amplification of the Hobbesian point about fallibility in the case of deep questions about morality and value rather than an endorsement of any sort of relativism. See Rawls. Thanks to Christopher Morris for emphasizing that the relativism Hobbes endorses is not a form of moral relativism, but rather a form of relativism about non-moral value.

³⁷Gauthier, "Public Reason," p. 36.

³⁸I suggest that this is not what he has in mind when he speaks of a theory of justice because the questions he is putting off on that theory seem to be questions about the substance of the principles of public reason, rather than questions about their scope. In particular, Gauthier notes that the standard in question must "provide some accommodation for each citizen's individual concerns" (Ibid., p. 36) just before noting that such issues belong to the "province of the theory

of justice." Moreover, as will become quite important later, Gauthier does *seem* to give us a hint of what he thinks such a theory would look like when he helps himself to the supposition that the citizens should see themselves as retaining a right to reconsider their authorization, "should they find their wishes and expectations betrayed." (Ibid., p. 37)

Finally, one might consider the theory of justice to be the theory that explains how we determine whether any particular individual would have been selected as the public person in a hypothetical contract. As I note in the text, this would only make it a theory of what the principles of public reason are to be like in a very indirect way, by specifying who gets to lay them down by fiat, and as I have just argued, Gauthier seems to have in mind something more direct and substantive than that. Indeed, he seems to have in mind a theory that would in fact constrain what the public person may justifiably do (he must accommodate "each citizen's individual concerns" to some degree, in particular), but as should become apparent, such a theory is at odds with his project.

³⁹See Gauthier, "Public Reason," p. 38. Here, I have essentially taken over Gauthier's formulation and made the relevant alterations.

⁴⁰Indeed, Gauthier's own brief comments about how we might determine whether a given public person is justified in acting as such seem to embody this point, when he argues that, "where equity is lacking, the actual public person would not have received the authorization of at least some of the citizens in any reasonable initial agreement." (Ibid., p. 40) One wants to ask, "Why not say instead that where equity is lacking the *conception of public reason being embodied by the prevailing institutions and practices* would not have received such authorization?"

⁴¹Precisely how such a code would be enshrined is a detail I shall not focus on in the present essay, though one could imagine the preamble of a public constitution serving this purpose, at

least in certain context. Moreover, it is not altogether obvious that the standard would need to be quite so explicit and enshrined; certainly, standards for evaluation in various contexts *can* emerge in a more gradual way and have enormous influence without ever being officially enshrined in some document. I assume some more officially sanctioned version of public reason for present purposes both for simplicity and because such an account more closely resembles Gauthier's own account.

⁴²I am not at this point assuming that the "public person" needs literally to be just one particular person rather than an assembly of some sort, though Gauthier is not particularly clear about this. Further, if the main objection of the present section is successful, one might worry that he could not consistently think of the public person as not literally being just one person. For a useful discussion of a very similar point in discussing Hobbes's not completely ruling out forms of government, see Hampton, pp. 105-107. What *is* relatively clear is that Gauthier is not using "public person" as a metaphor - he is very much interested in having some actual agent(s) define the standard of public reason, though that agent's being justified in so doing will be a function of the hypothetical agreement of the citizens.

⁴³Gauthier, "Public Reason," p. 39.

⁴⁴Here and throughout much of my exposition of Hobbes's "regress argument" I follow Hampton's useful account quite closely. See Hampton, pp. 101-102.

⁴⁵Hobbes, *Leviathan*, pp. 205-206.

⁴⁶Hobbes, *Leviathan*, p. 240.

⁴⁷Here I follow Hampton quite closely, see p. 98. Hampton also usefully discusses how this argument is more effective than many have supposed against the suggestion that we can have a system of government that is "closed" (has a single authority beyond which no appeal can be

made) without "there being a simple and indivisible power residing in a *person* or *assembly of persons*." (Hampton, p. 98)

Of course, Hobbes does allow that you may permissibly disobey the sovereign if he orders your execution, or if he commands you to incriminate yourself, and Hampton argues that this means Hobbes's own view is impaled upon the regress argument. Indeed, her case against Hobbes is in this way similar to my criticism of Gauthier's revised Hobbesian view.

⁴⁸Interestingly, Kant deployed just this argument in favor of absolute obedience to the sovereign:

...even the constitution cannot contain any article that would make it possible for there to be some authority in a state to resist the supreme commander in case he should violate the law of the constitution, and so limit him. For someone who is to limit the authority in a state must have even more power than he whom he limits, or at least as much power as he has; and, as a legitimate commander who directs subjects to resist, he must also be able to protect them and to render a judgment having rightful force in any case that comes up; consequently he has to be able to command resistance publicly. In that case, however, the supreme commander in a state is not the supreme commander...

Immanuel Kant, *The Metaphysics of Morals: The Doctrine of Right* (Cambridge: Cambridge University Press, 1991), p. 130. Kant also, like Hobbes, cannot fully stomach the consequences of this view, and sometimes allows that there is a right to disobey laws that directly contradict "inner morality." For useful discussion, see Thomas Pogge's "Kant's Theory of Justice," *Kant-Studien* 79 (1988), pp. 407-433. There are, of course, many important differences between Kant's account and Hobbes's; for Kant, the citizens do not regard the sovereign as their own, though they may not disobey them simply because they disagree with them. The main point for present purposes is that someone whose ultimate metaphysics of value are as radically different from

Hobbes's as Kant's takes a position on political authority that is *very* similar to Hobbes's (so similar, I submit, that the objection being lodged in the present essay against Hobbes would be quite forceful against Kant as well).

⁴⁹One might object that the difference between an arbitrator who simply fiats whatever standards he likes for public deliberation and one who must interpret some officially enshrined conception of public deliberation is an important one. Even if Hobbes is right that any written law, rule, or ideal is subject to widely divergent interpretations, however, this objection is not all that forceful. Of course, one might rightly worry that Hobbes has exaggerated the degree to which such official documents are subject to warped interpretations, and I ultimately want to endorse this worry. Let us grant Gauthier this point for the moment, though, since I shall argue that he cannot consistently deploy this line of argument in the first place.

⁵⁰Since Gauthier explicitly concedes that his formulation of the scope limitation in terms of matters that "significantly affect the interactions of citizens and the public goods available to them" is in his words "deplorably vague," it hardly bears repeating that his account as currently formulated is especially vulnerable to this charge. However, this would be unfair to Gauthier, since he only intends this formulation as a "first approximation to the appropriate scope limitation." Still, it does highlight how important it is for him to make good on this promissory note and provide a tighter version of the scope limitation if he wants to pursue the regress argument against the alternative account I have proposed.

⁵¹Gauthier, "Public Reason," p. 37.

⁵²Gauthier seems to want to disguise the fact that he is giving them this power rather than simply recognizing that they will inevitably claim it anyway by putting his point in terms of how it would be "futile to expect" them not to retain such a right. Even if he was only making this weaker

claim, however, insofar as the regress argument is plausible as put forward in its strong form by Hobbes, this would still doom Gauthier's account, as it would imply that no arbitrator will ever actually be able to retain power because his citizens will inevitably interpret these questions in self-serving ways whenever it suits their interests.

⁵³The discussion which follows in the text is heavily indebted to the comments of this anonymous editor. Since the point I make here is extremely helpful in clarifying my own account, I am quite indebted to this anonymous editor for bringing it to my attention.

⁵⁴Actually, we would also need an explanation of how the regress argument as posed against Gauthier's theory is unsound but I shall overlook this point.

⁵⁵Here I make the assumption that for any theory of public reason to count as Hobbesian it must make the authorization of a public person primary and the standards of public reason derivative. Perhaps, however, one could construct a theory of public reason that was recognizably Hobbesian and did not do this. If this assumption is implausible, then everything I say could easily be reformulated in terms of that subclass of Hobbesian theories that does give authorization of the public person such primacy.

⁵⁶Roughly and in brief, the connection is this. The present objection involves presenting any Hobbesian theory of public reason with a dilemma, and the Hobbesian faces one of these horns for essentially the same reasons that Gauthier cannot consistently use the regress argument.

⁵⁷Here, of course, I have no illusions about the novelty of my argument. We might call this horn of the dilemma the "Lockean horn," as John Locke famously made this point against Hobbes. Locke argued that the parties in the state of nature would never rationally grant the sovereign absolute arbitrary power, as this, "were to put themselves in a worse condition than the state of nature wherein they had a liberty to defend their right against the injuries of others and were upon

equal terms of force to maintain it, whether invaded by a single man or many in combination. Whereas, by supposing they have given up themselves to the absolute arbitrary power and will of a legislator, they have disarmed themselves and armed him to make a prey of them when he pleases, he being in a much worse condition who is exposed to the arbitrary power of one man who has the command of 100,000, than he that is exposed to the arbitrary power of 100,000 single men, nobody being secure that his will, who has such a command, is better than that of other men, though his force be 100,000 times stronger." John Locke, Two Treatises of Government, ed. Thomas Cook (New York: Hafner Publishing, 1956), p. 191. Of course, Locke essentially tries to evade the other horn of the dilemma by arguing that individuals in the state of nature, relying on their natural reason, could discern certain natural rights and duties, challenging core Hobbesian assumptions about the limitations of natural reason that I am conceding for present purposes (as well as running afoul of Gauthier's constraint that we see the world as "normatively disenchanting").

⁵⁸Here I bypass Hobbes's own alternative of having God enforce standards upon the sovereign as implausible and no longer helpful.

⁵⁹In particular, one might follow Rawls in supposing it to be misguided to try to derive the reasonable from the rational.

⁶⁰Fred D'Agostino has usefully categorized theories of public reason along three dimensions, and suggests that Gauthier's account is "convergentist, volitional, and pragmatic," while Rawls's account, by contrast, is "consensual, cognitive, and pragmatic." See his "The Idea and the Ideal of Public Justification," Social Theory and Practice, 18 (1992), 143-164 at 146. D'Agostino suggests that Gauthier's theory differs from Rawls's in giving different weight to the four desiderata that he argues motivate our evaluations of various theories of public reason. One

fruitful way of reading the present essay is as an attempt to argue that even if one gives the weightings and priorities to these desiderata that Gauthier's approach tacitly does (and hence try to derive the reasonable from the rational), you do not end up with Gauthier's theory. Hence, switching to the Rawlsian approach would be to accept a differing set of weightings and give up on the attempt to derive the reasonable from the rational, and would be unhelpful in the present context.

⁶¹Recall that if they rely on public reason there will be *no* limit on the sovereign since, on Gauthier's account, her will *defines* public reason.

⁶²Hampton has argued that democratic regimes are especially well-suited to solving Hobbesian conflict through public principles and the rule of law. They control disagreements by channeling them "into peaceful political paths" rather than relying on "a sovereign to banish such disagreement." (Jean Hampton, "Democracy and the Rule of Law," Nomos XXXVI: The Rule of Law, Ian Shapiro, ed. (New York: New York University Press, 1994), p. 40) Hampton's view is similar to mine, but her account does not rely upon the Gauthier-style transformation of our natural reason into public reason that my account does. This allows her to avoid controversies surrounding the notion of such a transformation, but it *may* also limit her results to the case of democratic regimes. Since she is, in the essay just quoted anyway, primarily interested in a theory of how democracies are especially well positioned to blunt the force of the regress argument, this is no vice, and probably a virtue, of her account. My aims in the present essay, however, are broader; I am trying to illuminate a strategy for generating theories of public reason that will apply even in contexts where non-democratic forms of government or perhaps pre-statist conventions (as in hunter-gather societies, for example) are justified. Thanks to Christopher Morris for bringing this element of Hampton's work to my attention.

⁶³My understanding of the G/wi is drawn from George Silberbauer's "Ethics in small-scale societies," chapter two of Peter Singer's A Companion to Ethics (Basil Blackwell, Oxford: 1991), 14-28 at 14-28. My choice of a small-scale society is not a coincidence, as on the face of it one might suppose that such societies lack what Rawls calls reasonable pluralism (indeed, they seem to lack even pluralism simpliciter, though this is difficult to determine from the outside - what might seem like a minor difference of opinion to an outsider might seem like a very deep, irreconcilable disagreement from within the culture), as well as a democratic background culture. So even if the Rawlsian approach is more fruitful when these (and other) special conditions obtain, it would not be particularly useful in assessing the conception of public reason we find in such small-scale societies (nor would Rawls claim that it would be). I mention this only to emphasize that (a) I am not reconceptualizing the problem being faced, which would simply be to change the subject, and, more importantly, that (b) which strategy for generating a conception of public reason is the appropriate strategy may vary from one context to another.

⁶⁴Again, following Gauthier, I leave aside the interesting question of how this transformation of one's reason within a given domain is supposed to occur. It is also probably worth noting that in small-scale societies like the G/wi it is quite likely that the public/private distinction may collapse or be deeply blurred, as anonymity and privacy are much harder to come by in such societies, and that as a result there may be little or no room (no "private sphere") for "natural reason" to operate at all. For present purposes, I will pass over such complications.

⁶⁵Whether such controversies actually arise and whether they would provoke any real controversy is not something I am in any position to speculate about. Here I am simply stipulating certain propositions about a hypothetical society much like the G/wi to make a philosophical point, and

do *not* mean to be doing arm-chair anthropology. Indeed, I imagine it is quite implausible as a piece of anthropology.

⁶⁶Here, of course, my proposal is quite similar to (perhaps an instance of) the sort of reflective equilibrium methodology made famous by Rawls.

⁶⁷It is worth noting that Joseph Raz has worked out a clear and elaborate account of such defeating conditions, though a discussion of his account would exceed the scope of the present essay. See, for example, his Practical Reason and Norms (London: Huteson and Co., 1975). Thanks to Christopher Morris and Gerald Postema for emphasizing the connection to Raz's position.

⁶⁸This becomes less and less plausible the further the dominant conception of public reason is from the ideal for the society in question, as the psychological strain of resisting a slide back into one's natural reason may be irresistible in contexts with conceptions of public reason that fall very short of the ideal.

⁶⁹Indeed, more generally, it is hard to see how the proposed alternative could be any *less* plausible than the Hobbesian one, especially given the vulnerability of Hobbes's account to the "Lockean horn" and the modified Hobbesian's inability to press the regress argument against the alternative account. I of course maintain that it is a significant improvement over the Hobbesian account. It is worth noting, however, that even if it were in the end only equally as plausible as its Hobbesian competitors that it would still provide those with Hobbesian sympathies a plausible alternative conception of public reason.

⁷⁰It is important to remember that the alternative I am suggesting is really a different strategy for generating a theory of public reason, which is consistent with one's being able to generate many different and perhaps plausible theories using that strategy. The theory I have been sketching in

the present paper is merely meant to be illustrative of how that strategy might be employed to generate such theories, rather than as a fully developed theory that I am endorsing.

⁷¹Indeed, such a modification would seem to impale Gauthier's theory on the first, Lockean horn of the Hobbesian Dilemma.

⁷²Christine Korsgaard, for example, has recently leaned heavily on Wittgensteinian considerations in providing her own distinctively Kantian account of the reasonable. See her [The Sources of Normativity](#) (Cambridge: Cambridge University Press, 1996).