The Coherence of Public Reason

1.   Introduction

Of the many criticisms aimed at theories of public reason, three of the most fundamental are that these theories are self-defeating, that they lead to anarchy, and that they treat those classified as unreasonable with unequal respect. Self-defeat is a problem for any theory, while anarchy and unequal treatment are problems for conceptions of public reason intended to be compatible with liberal-democracy, which in practice means all of them. All three criticisms suggest that public reason is incoherent, rather than simply outweighed by competing considerations. To avoid anarchy and self-defeat, advocates of public reason can idealize heavily, adopting a strict standard of reasonableness. Yet the more they restrict the justificatory constituency, the more the resulting exclusion of putatively unreasonable views seems unfair. Theories of public reason founder on the shoals of anarchy and self-refutation, on the one hand, or unequal treatment on the other. The purpose of this paper is to show how we can avoid this dilemma.

Worries about self-defeat, anarchy, and the unfairness of exclusion are not new.[1] David Enoch has recently restated them with great clarity, however, and at a level of generality intended to undermine not just this or that particular theory of public reason, but all such theories.[2] I will therefore focus on his presentation of the objections. Although he doesn't explicitly identify the dilemma sketched above, it is consistent with his view to recognize that the solution to the first two problems exacerbates the third.

My suggestion about how to avoid this dilemma depends on a distinction between two models of public reason. The first frames the principle as a constraint on coercive state action, with a default of inaction; the second sees the principle as a constraint on reasons for decisions, with a default of exclusion from consideration. The coercion model is vulnerable to the worry about anarchy but not self-defeat, I argue, the reasons model to the worry about self-defeat but not anarchy. The coercion model does not apply to itself, and so cannot be self-defeating. In addition, there may be ways of avoiding anarchy without idealizing heavily. As a result, this model is less vulnerable to the charge of unfair exclusion. I favor the reasons for decisions model, however. This model is not subject to the anarchy objection, but does apply to itself. The only way to avoid self-defeat, given that this version of the principle is reflexive, is to idealize heavily, declaring all those who don't accept public reason to be unreasonable. The reasons model is thus particularly susceptible to the worry about unequal treatment and unfair exclusion. My response is to appeal to reciprocity, in order to justify making acceptance of public reason one of the conditions of reasonableness (thus protecting the view from self-defeat). Those who reject public reason exclude everyone but themselves from the constituency of justification, and so are not well placed to object that others draw the boundaries of this constituency too narrowly.

2.   A dilemma for public reason liberalism

The general idea of public reason is that the exercise of political power should be justifiable to all concerned, in the sense of being acceptable to each despite the fact that people reasonably espouse different religious or philosophical doctrines. As the reference to reasonableness suggests, the constituency within which unanimity is required is restricted and/or idealized, though not so much as to eliminate perspectival diversity, yielding what David Estlund refers to as a "qualified acceptability requirement."[3] According to David Enoch, the motive for such idealization is to solve a problem created by the way public reason liberals try to reconcile

individual liberty with political authority (114-15, 117). If genuine consent to membership in a political society were necessary for legitimate political authority, such authority would be rare, because as Hume pointed out the poor unilingual peasant has no realistic exit options.[4] Hume concluded that the justification of political authority must be its social utility; philosophical anarchists conclude there is no legitimate political authority. According to Enoch, public reason liberals take a third path. Political authority need not be freely accepted by all given some genuine alternative, but only justifiable to all. What matters is not whether everyone had a real choice about joining a political society, but whether the character of the regime is such that everyone has reason to accept it. Unanimous acceptability cannot involve actual acceptance on the part of every person, however, for the diversity of conflicting doctrines accepted by citizens would mean that nothing at all would pass the test; "[a]narchism follows" (118). So too does self-defeat, if the principle applies to itself.[5] We must idealize; public justifiability can only require acceptability to a limited set of perspectives that satisfy some further standards.

Enoch distinguishes two "idealization devices": exclusion of unqualified points of view, and "going hypothetical" (117), i.e. requiring acceptability not to people as they are, but as they would be if placed in ideal conditions, e.g. full information (127). I will assume that both kinds of idealization are necessary. It might be that even under ideal conditions some people would still subscribe to frankly illiberal, undemocratic doctrines. Conversely, even if we have narrowed the range of qualified perspectives by requiring acceptance of certain basic moral beliefs, we must still idealize conditions of deliberation, because espousing a reasonable doctrine does not guarantee that one will draw reasonable conclusions from this doctrine, if one is not in good epistemic conditions, or lacks adequate cognitive capacities.

In everyday language, I am reasonable if I am willing to moderate my claims in response to the claims of others, consider issues from other people's perspective, and play by the rules

when others are likewise playing by the rules. Nazis and psychopaths fail this standard. But the standard of reasonableness invoked by many public reason theorists is much higher. Enoch cites Burton Dreben, David Estlund, and Jonathan Quong on the fact that one of the criteria for counting as reasonable is acceptance of the principle of public justification (121, note 26). Even seemingly reasonable liberals such as John Stuart Mill, Jean Hampton, and Joseph Raz therefore count as unreasonable, according to these accounts.[6] Yet public reason theorists cannot retreat to the everyday conception of reasonableness, Enoch contends, because "the price would be anarchism," there being nothing that is justifiable to all those who are reasonable in the ordinary sense (122).

Although he does not say so explicitly, Enoch has identified a dilemma for public reason liberals. The more we idealize in order to avoid anarchy and to ensure that all those deemed qualified accept the principle of public justifiability, the more it will seem unfair to exclude the putatively unreasonable from the constituency of justification. Conversely, if we exclude only the manifestly unreasonable, so as to weaken the charge of unfairness, little if anything will be justified, making the principle inconsistent with support for liberal-democratic institutions, and making it controversial amongst the set of points of view deemed qualified, leading to self-defeat.

This dilemma presents a serious challenge for theories of public reason. I believe it can be avoided, if we distinguish two models of public reason. In the next section I present these two models, explaining how anarchy is a danger for one, self-refutation a danger for the other.

3.   Two models of public reason[7]

According to what I will call the reasons-for-decisions model of public justification, the reasons that lie behind our political decisions must pass the qualified acceptability test, otherwise we exclude the reasons in question from our deliberations. By "political decisions," I mean

decisions about coercion in the context of a less-than-fully-voluntary association claiming final say about such matters. Whether the principle of public reason applies to all political decisions or just some class of fundamental political decisions is a question I will not consider, since it is orthogonal to the issue that concerns me, which is whether the principle is framed by a default of exclusion or inaction. On the reasons view, the default is exclusion, with a presumption against including a reason in one's calculation of where the balance of reasons lies, a presumption defeasible only by unanimous acceptability amongst qualified points of view. I interpret Rawls' account as being based primarily on the reasons model. According to what I will call the coercion model of public justification, in contrast, the laws themselves must pass the qualified acceptability test, otherwise we refuse to enact any law, in the domain in question. On the coercion view, the default is inaction – not having any enforced rule, on the matter at hand, generating a presumption against state action defeasible only by unanimous qualified acceptability.[8] I associate the coercion view primarily with the work of Gerald Gaus and Kevin Vallier,[9] although the two models are meant to be pure or ideal types, not complete descriptions of any one person's theory of public reason.

The coercion view requires unanimity about a proposed law based on each qualified point of view's total balance of reasons, rather than on the balance of shared reasons.[10] Since each doctrine accepts some different reasons, the coercion view demands convergence from different perspectives. In contrast, the reasons view requires unanimity about reasons, while permitting disagreement about where the balance of public reasons lies. The fact that one requirement can be satisfied without the other shows that the two are not equivalent. On the one hand, it is conceivable that all reasonable or otherwise qualified points of view would support a given law but for different reasons, even while the balance of public reasons was negative. For example, even if the balance of reasons that feminists and conservatives share does not justify restricting

pornography, it might be that each perspective's total balance of reasons does. The coercion view is less restrictive of state action in this respect, since it permits non-public reasons to justify state action opposed by the balance of public reasons. On the other hand, it is likely that in certain cases the balance of reasons that all qualified points of view accept will be positive, but that some qualified point of view will object to the proposal based on the full set of reasons it accepts. For example, the shared reason of economic prosperity might speak in favor of developing a territory that happens to be of special spiritual significance to members of a particular religion. The coercion view is more restrictive in this respect, since it permits non-public reasons to block state action supported by the balance of public reasons. The two frames could also be combined in various ways, yielding more complicated, hybrid models.[11]

For the coercion view, anarchy is the main danger, because this model requires unanimity in order to overcome the presumption against state action. True, there is the possibility of unanimous convergence from different perspectives even if the balance of shared reasons is negative, as in the pornography example. However, in that case there were only two perspectives: conservatism and one kind of feminism. Unless the set of qualified perspectives is severely restricted, anarchy looms. Anarchy is not on the horizon, however, if public justification simply constrains our reasons for political decision-making. The lower the threshold of qualification, the more diversity there will be in the set of qualified perspectives, and the fewer reasons will count as public. Fewer public reasons means a simpler cognitive task, in deciding where the balance of public reasons lies, not a presumption against enforcing common rules. The main danger for the reasons model is self-refutation. The principle of public reason is itself a reason that one might need to invoke in political decision-making in order to rule out non-public reasons and thus shift the balance of admissible reasons one way or the other. In contrast, according to the coercion model there is no constraint on reasons for political decisions, and

therefore no way the principle can apply to itself.

4.   Why the coercion model avoids self-refutation, and may not require heavy idealization in order to avoid anarchy, and so is less vulnerable to the charge of unfair exclusion

The idea that public reason might be self-defeating arises because the principle enjoins avoidance of controversy and is itself controversial. As Enoch puts it, "the main ideas underlying the public reason tradition... are just as controversial as whatever first-order views public-reason theorists are trying... to abstract from."[12] In order to be self-defeating, disagreement-avoidance principles must meet two conditions. First, *reflexivity*: does the principle apply to itself? If not, failure to pass its own test will not matter. Second, *controversy*: is the principle controversial, within the relevant population? If not, the fact that a principle applies to itself will not matter. Self-defeat can therefore be avoided in two ways: by showing that the principle is not reflexive, or that it is not controversial.[13] Enoch claims that there is no rationale for either of these responses except for saving the theory from self-defeat. In this section I focus on the issue of reflexivity, because I claim that the coercion model is not reflexive; in the next section I focus on controversy, because the reasons model *is* reflexive, and so has to made non-controversial (and can be, I claim, via appeal to reciprocity).

It is possible to make public reason non-reflexive by stipulation, e.g. by adding to the principle the introductory clause "except for the present principle…" This addition will be ad hoc, however, unless it has an independent rationale. If the case for public reason rests on reconciling liberty with authority, and this requires reasonable acceptability of the principles on the basis of which authority is exercised, then the principle would seem to apply to itself, because it is among the reasons on the basis of which authority is exercised. On the coercion model, however, the principle of public justification does not apply to the principles upon which

political power is exercised, but to the exercise of political power directly. So long as my doctrine is reasonable or qualified, I can object to a law based on my doctrine's total balance of reasons, and that objection ought to be decisive, even though it draws upon non-public reasons. Conversely, even if the balance of shared reasons is negative, unanimity in favor of the law based on each doctrine's total balance of reasons would make the law legitimate, even though people would be supporting the law for (different) non-public reasons. On this view, reasons for political decisions need not be unanimously acceptable amongst qualified points of view. Because the coercion interpretation of public reason applies at the level of laws, with a default of collective inaction, and not at the level of reasons, it cannot apply to itself, and therefore cannot refute or exclude itself. Controversy about the principle is therefore no danger to its validity, on this account.

In his response to the self-defeat objection, Gaus rejects reflexivity by drawing an analogy between public justifiability and falsifiability. The principle of falsifiability need not be falsifiable because it is not a scientific argument; rather, it is a meta-claim in the philosophy of science about what constitutes scientific argument.[14] This appeal to different levels of discourse is unnecessary, however, if what requires public justifiability is not reasons for or against coercion but coercion itself. On the coercion model, the principle of public reason would be reflexive only if the principle were to be coercively enforced – if, for example, those uttering non-public reasons in the context of political debate were to be fined or imprisoned, or if those carrying placards with non-public slogans were to be arrested. If the principle is simply part of an ethics of citizenship, however, that is not the case. Criticizing someone for supporting publicly unjustifiable coercion is not itself a form of coercion.

Enoch's claim that the rationale for public reason makes the principle reflexive is most plausible for Gaus' recent work, where the demand for public justifiability applies in the first

instance to claims of authority, claims that are allegedly implicit in ordinary social morality. If moral claims involve people claiming authority over one another, then when we criticize others for not following public reason we will be exercising authority, making the principle reflexive.[15] As Enoch points out, however, there are reasons to doubt that moral demands involve claims to authority. When I say "you should teach your daughter to read!", to use Enoch's example, I am not claiming that you must submit to my authority, but to the authority of morality.[16] Gaus' response would be that if I persist in issuing this demand even after you have explained your reasons for disagreeing, then I am demanding that you defer to my judgment. In face of such disagreement, however, I may stop making demands until such time as the matter is settled politically. In short, we can think of moral demands as involving either rational argument about what is right, or appeals to *political* authority, neither of which involves any demand for deference to personal judgment. On this interpretation, the coercion model is not reflexive, and therefore cannot be self-defeating.

The worry about anarchy remains, however. The looser our standard of qualification, the broader the range of perspectival diversity will be, making it less likely that all perspectives will agree which policy is optimal. Defenders of the coercion model can respond by invoking "nested inconclusiveness" or "higher-order unanimity."[17] Qualified perspectives will disagree about what policy is best, but may agree that a range of different policies are better than nothing. A law will therefore be legitimate, even if many think it less than optimal, so long as it is selected by some reasonable procedure from the set of proposals that are unanimously preferred to the option of not having any common policy. This set of policies Gaus dubs the 'optimal eligible set.'[18] The model will still lean libertarian or anarchist if we apply the qualified acceptability criterion to laws at a highly disaggregated or finely individuated level. It might be true that for each narrowly delimited issue, there is some qualified point of view that prefers the option of having

no common rule, assuming a range of other policies is in place. If we package issues together, it is less likely that there will be a qualified point of view that rejects having any system of common rules at all, for the bundle of issues at hand. No doubt there are some systems of law that are worse than anarchy. But there is a range of legal and political regimes that, taken as a package, it would be unreasonable to reject, as against the alternative of a state of nature. Thus, aggregation of issues into bundles (and use of decision-procedures to select from bundles agreed to be better-than-nothing) may allow weak or moderate standards of qualification to avoid anarchy.

This strategy will seem ad hoc unless we have a reason for aggregating issues other than avoiding implausible conclusions, a reason that follows from the more fundamental concerns public reason is meant to address.[19] Gaus' proposal is to consider issues independent unless some member of the qualified public cannot express a preference at that level of aggregation. Suppose I reasonably believe that private property is justified only if accompanied by a welfare state, and that otherwise it would be morally preferable to have no enforced scheme of property rights at all. How are we to register such a preference, if we are trying to apply an idealized unanimity standard to the decision about property rights *independent* of the issue of redistribution? When I am asked whether I approve of private property rights, I cannot give an unambiguous answer; I can only say "yes, if accompanied by welfare state policies, no otherwise." If we are demanding acceptability to all qualified points of view, we cannot pitch our demand at a level of disaggregation such that some qualified points of view cannot express a preference. We must aggregate, which in this case means considering whole packages of property rights and welfare policies. In contrast, no one can reasonably think 'I am in favour of state promotion of high culture, and if I can't have that, then I prefer that we get rid of property rights too.' There is nothing obviously unprincipled about Gaus' standard. The most holistic reasonable view may

force a lot of aggregation, undoing some of the classical liberal tilt he thinks results from the

demand for public justifiability, but that doesn't make the criterion unprincipled.

5.   Why the reasons model avoids anarchy, but needs heavy idealization in order to

avoid self-refutation, and so makes itself especially vulnerable to the charge of

unfair exclusion

I said earlier that framing the principle as a constraint on reasons for political decisions

avoids the anarchy objection, because this model involves no presumption against state action.

The requirement is that decisions about whether / how to act be taken based on public reasons,

not that we act only if no one can reasonably object (based on their total balance of reasons).

There is a price to pay for framing the principle of public justification in this way, however,

which is that it makes the principle reflexive, and so potentially self-defeating. When framed as a

constraint on reasons-for-decisions the principle is at risk of self-refutation, because it is itself a

reason invoked in political decision-making. This is why Estlund claims that a qualified

acceptability requirement must apply to itself.[22] If public reason is framed as a constraint on

reasons for decisions, the only way to avoid self-defeat will be to claim that the principle is not

controversial within the relevant constituency.

One way to establish that the principle is uncontroversial would be to provide a criterion

of qualification, and then to show that, as a matter of empirical fact or logical necessity, all

points of view meeting that criterion do or must accept the idea of public reason. For example,

one might define 'reasonableness' as (a) the ability and willingness to reason with other people

about what is true and good and right; (b) recognition of the existence of obstacles to

understanding and agreement, on the part of people satisfying (a), and; (c) acceptance of the idea

of society as a fair scheme of cooperation between free and equal persons. It seems doubtful,

however, that everyone who is reasonable in this sense accepts or must accept the idea of public

reason. The most prominent defense of the principle rests on the value of equal respect for persons as rational and moral agents. As Steven Wall argues, however, it is not unreasonable to deny that equal respect requires this kind of multi-perspectival acceptability, because it is at least plausible that equal respect requires that we treat people in ways that are truly justified.[23]

The other way to make public reason uncontroversial is to claim that acceptance of public reason is itself one of the criteria of qualification. In order to count as fully reasonable, one must accept the principle of public justification, making the principle uncontroversial by fiat.[24] However, unless we have some reason for declaring all non-public-reason political philosophers unreasonable, this move will seem like "an ad hoc restriction introduced just in order to save the theory" (123). Moreover, the restrictiveness of this definition of the reasonable makes the unfairness of excluding the unreasonable all the more apparent, since this group will include not only Nazis and psychopaths but good liberals who simply reject the principle of public reason.

There is, however, an independent ground for thinking that the reasons for which we make political decisions need to be acceptable only to those who accept public reason, which is that the principle involves accepting moral costs for the sake of having the appropriate relations with others. According to its advocates, the point of the principle is to make possible a relationship of civic friendship across deep disagreement.[25] If the justification of public reason is relational, in this way, it is natural that the principle should involve a reciprocity condition. The question of public justifiability arises with respect to reasons that I think are valid or sound, and which would be politically relevant, but for the fact that they are reasonably rejectable. The principle of public justification asks me to exclude these reasons from the bases of my political conduct despite their being in my view true. This exclusion of reasons comes at some moral cost, judged from my comprehensive point of view. Given what is being asked of me, it is reasonable that I should be willing to exercise restraint only with respect to perspectives that themselves

recognize a duty of restraint. The reason I have to care about acceptability in regard to other points of view does not apply with respect to points of view that do not care about qualified acceptability.

Enoch is skeptical about the appeal to reciprocity because at the most fundamental moral level reciprocity does not matter. "I don't recall, in Kant, the proviso that what we should always treat as an end and not merely as a means is only the humanity of those who would treat us in a similar way" (124). I grant that many basic moral rights and duties are not conditional on reciprocity. The duties not to torture or murder people are not suspended with respect to known torturers or murderers. For other moral rules, however, a reciprocity condition makes more sense. With respect to distributive justice, for example, if one subscribes to a relational conception of egalitarianism, one will find it plausible that the concern with distributive equality applies only where there is some assurance of reciprocity.[27] Similarly, if the point of public reason is to create relations of friendship despite disagreement among citizens, it seems natural to suppose that the duties associated with public reason are not unilateral, but must incorporate a reciprocity condition.[28]

One complication in thinking about reciprocity and public reason is that we are never dealing with relationships between only two parties. Some people are committed to making decisions on public grounds while others are not, but laws apply to everyone. Am I entitled to ignore public reason because some people are not committed to the principle, or does my relationship with those who are committed to the principle take priority? We can distinguish two forms a reciprocity condition may take. In the first, I am prepared to abide by public reason so long as some minimum percentage of my fellow citizens is likewise willing. In the second, I am committed to abiding by public reason regardless of the popularity of the principle, but I take the principle to exclude from political decision-making only those reasons that are controversial

amongst points of view accepting the principle of public reason. The first form of the reciprocity condition makes sense if we view the principle of public reason as one we adopt in view of the expected effects of its regulating our behavior, since these effects may depend on the overall level of compliance. The second form makes more sense, however, if we have a relational conception of public reason, according to which acceptance of the principle affects the nature of our relationships. Suppose A and B accept the principle but C does not. The fact that A cannot have the right kind of relationship with C (one of civic friendship) does not negate the importance of A and B having such a relationship. In my view, we should think of the principle of public reason as being binding with respect to our exercise of political power over everyone, and conditional on reciprocity in that it requires acceptability only in regard to perspectives accepting the principle.

The issue of the form of and rationale for the reciprocity condition obviously require further discussion.[29] What I hope to have established here is just that it is plausible that the principle should be subject to such a condition, if its justification is relational, and that in this case the self-refutation objection would be met.

6.   The unfairness of exclusion objection

So far, we've seen that the coercion model is vulnerable to the anarchy objection, but may be able to avoid it via higher-order unanimity and nested inconclusiveness. I prefer the reasons model, which avoids the anarchy objection but is vulnerable to self-defeat. It can avoid self-defeat by appealing to reciprocity. In doing so, however, the reasons model idealizes heavily, and thus exposes itself to a particularly sharp form of the objection from unequal treatment and unfair exclusion.

In one form, the objection that public reason involves unequal treatment is question-begging. If the objection is that the principle imposes unequal costs on people as compared to the

alternative scenario in which no such principle applies, the objection can be met simply by disputing this baseline of comparison. As Robert Nozick argued, it does not count against laws against rape that they have a differential impact on men and women as compared to a state of nature, because the law in question is independently justified.[31] Differential impact as compared to a baseline in which people are permitted to violate each other's rights and are differentially inclined to do so is not unfair. Defenders of public reason claim that exercising political power on the basis of non-public reasons is wrong. If they are right, it is no objection that public justification places heavier demands on some than others, compared to the situation in which no such principle prevails.

The objection based on unequal treatment / unfair exclusion can be reformulated, however, so as to avoid this response. The problem is not differential impact relative to the no-principle scenario, but differential treatment inherent in the principle itself. It's true, in one sense, that the principle treats everyone the same, since it enjoins everyone from making political decisions on reasonably rejectable grounds (or, on the coercion model, from supporting laws that some reasonable point of view rejects, as opposed to not having a law at all). As these formulas suggest, however, the principle employs a distinction between reasonable and unreasonable points of view. Joseph Raz claimed that this distinction involves a form of differential treatment that is inconsistent with public reason's animating values.

> It is not evident why if legitimacy does depend on agreement it should depend
> only on the agreement of those with reasonable views. A prima facie objection to
> restricting the required consent in that way is that every person counts. The life
> and well-being of those with unreasonable views are just as likely to be affected
> by the actions of political authorities as the life and well-being of other people.
> Moreover, their life and well-being are of moral consequence. They cannot be

ignored, and if other people's agreement is required, so should theirs be.[32]
Enoch makes the main point vivid by pointing out that "the unreasonable too are born free"
(123). The problem of reconciling liberty with political authority arises for each and every
person, reasonable or not. The solution is supposed to be never to coerce someone based on a
principle that they do not endorse. If we exclude the unreasonable from the scope of this demand
simply in order to avoid anarchy, we fail to treat them as equals. Public reason liberals cannot
exclude the unreasonable from the constituency of justification without straying from the core
commitments of public reason liberalism, tolerating differential treatment that should on their
own view be intolerable.

Part of the answer to this objection is just to clarify the extent of the differential treatment
at stake. Exclusion from the constituency of justification does not mean that the unreasonable are
denied the ordinary rights of citizenship.[33] The unreasonable cannot be denied the vote or the
right to free speech simply for being unreasonable – hate speech is another question – but
characterizing a view as unreasonable does not imply that it should be banned. Nor does this
exclusion deny that the unreasonable are owed a justification for the laws to which they are
subject. Like everyone else, the unreasonable have a right to know what the law is, why it applies
to them, and what its justification is. What they do not have is a normative veto over the reasons
for which we exercise political power (or, according to the coercion model, over the enforcement
of a common rule, in a particular domain). The life and well-being of the unreasonable are
undeniably important, but the exclusion of a point of view from the set of qualified points of
view does not deny this fact. The question is whether there is a form of morally objectionable,
substantively unequal treatment in my deciding how to exercise my own political rights
according to a principle of multi-perspectival acceptability that includes as qualified only *some*
perspectives. (Alternately, if one prefers to frame the principle as a constraint on state action

directly rather than on reasons for decisions, the question would be whether there is a form of substantively unequal treatment in permitting laws rejected by unreasonable views, but not those rejected by reasonable views.)

The second part of the answer is to make clear that there are different dimensions to the standard of qualification, and that a perspective can fail in one respect but not the others. To call David Enoch unreasonable because he rejects public reason is not to suggest that he is a fanatic or a bigot. A point of view can be unqualified without it or its supporters being unreasonable in the ordinary, non-technical sense (which is a merit of Estlund's generic term "qualification"). Exclusion from the constituency of justification does not mean exclusion from the ordinary rights of citizenship, nor necessarily express a judgment of unreasonableness in the ordinary sense.

The principle does still involve a form of differential treatment, however, and it may seem that even this difference in treatment is objectionable, from the perspective of public reason's animating values. Since everyone is born free and equal, the problem of reconciling political authority with liberty and equality arises with respect to every person, reasonable or not. If we water down this standard by restricting the constituency of justification solely in order to avoid anarchy or self-refutation, we may be by our own standards guilty of disrespectful, unequal treatment of the unreasonable. However, public reason liberals should simply deny the premise of the objection, that the exercise of political authority ought ideally to meet a standard of unrestricted unanimous acceptability.

I grant that it is one of the fundamental premises of modern political thought, found in the doctrines of Hobbes and Locke, for example, that no one can be subject to political authority without their own consent. This principle does suggest that a kind of unanimity is necessary for legitimate political authority.[34,35] However, Hobbes and Locke also accepted that it was

legitimate to use force in defense of one's own rights (and, in Locke's case, to protect the rights of others) without regard to consent.[36] Exercising this kind of power was justified simply by the necessity of enforcing the basic moral principles that constitute the law of nature. The origin of the principle of public reason is not to be found in the idea that political authority requires consent, but in recognition of reasonable disagreement about the law of nature, i.e. reasonable disagreement about the moral law setting out the rights and duties people take themselves to be justified in enforcing *without* regard to consent. Such disagreement threatens to undermine relations of mutual respect and civic friendship, given that membership in political society cannot be fully voluntary, on the model of the associations that exist within political society. The ideal of public reason asks me not to make political decisions on the basis of all of the reasons I think are valid or true, and it does so on the ground that some other people disagree with me. But why should I care about *that*, given that I think my views are correct and theirs wrong? The answer must involve my recognizing something like the following: despite our deep disagreements they, like me, are concerned about truth and morality and justice; they, like me, are doing their best to figure out what these ideals require, and are willing to deliberate sincerely with others about these questions; they understand society to be a fair scheme of cooperation between people who for reasons other than malice, self-interest, or intellectual negligence subscribe to different religious and philosophical doctrines; and (I would argue, though this is more controversial) that they, like me, are willing to accept the demand for multi-perspectival acceptability involved in the principle of public reason. If we did not restrict the range of perspectives whose acceptance is necessary for legitimacy, the moral costs[37] of the restraint the principle demands of me would be higher, and in many cases I would be paying these costs for the sake of friendship with people unwilling to enter into the relationship in question. Idealization is not an ad hoc solution to the implausible consequences of an unrestricted

acceptability condition that is in itself attractive. It is an essential feature of the moral problem to which public reason is a response, a necessary condition for a principle of public justification to have any moral claim on anyone, before we consider its potential consequences.

The claim that the reason for idealization is avoiding anarchy assumes that public reason liberals think that, consequences aside, there is something to be said for unrestricted unanimous acceptability, as a constraint on the legitimate exercise of political power – not just that it would be a good thing if everyone happened to agree, but that unless political authority can win the approval of all points of view, even the frankly unreasonable ones, this is a strike against the regime, a pro tanto bad. That strikes me as an implausible view, which public reason liberals should simply deny. Although the exercise of political power over anyone has to be justified, and justified to them in the ordinary sense of having its rationale explained and defended, there is nothing morally bad about exercising political power in ways Nazis and psychopaths cannot accept. Indeed, if we respect each person's moral freedom and equality, we must exercise political power in ways Nazis and psychopaths will object to.

I said above that reasonable moral disagreement poses a moral problem for the exercise of political power because a political society is not and cannot be a fully voluntary association, on the model of the various clubs and groups that exist within a political society. If we accept voluntariness as an ideal, however, does this not imply that there is something attractive about unrestricted acceptability? No. Even if it would be better if political societies could be made into genuinely voluntary associations, it does not follow that the rules of an unavoidably involuntary association ought to be acceptable to all people, no matter their moral views. If Nazis could go off and live by themselves, that might be a good thing, but it gives us no reason to think that the laws of an actual political society ought to be acceptable to them, since they do not believe in liberty and voluntariness. The belief that membership in political society would ideally be fully

voluntary suggests that laws ought *not* be acceptable to those who reject basic liberal ideals.

The case of liberals who count as unreasonable only because they reject public reason (e.g. Enoch, Wall, etc.) obviously cannot be dealt with in the same way. It is indeed unfortunate that we must exercise political power for reasons they cannot accept. Yet given that such liberals are not themselves prepared to exercise restraint, we do them no wrong, and we do not treat them unfairly, in refusing to count their points of view as qualified, for the purposes of defining the contents of public reason. This is the point about reciprocity, discussed above. Critics of public reason are poorly placed to complain that public reason liberals draw the circle of qualification too narrowly, unfairly excluding people from the constituency of justification, since they in effect draw the boundaries of qualification more narrowly still; correctness justification is simply public justification with the circle of qualification reduced to one. 'But I reject the whole idea of public justification,' Enoch might respond, 'so I'm not guilty of any inconsistency.' Neither are public reason liberals guilty of inconsistency, however, if the justification of the principle is relational in the way I have described, because rejecting public reason justifies the exclusion of one's point of view from the set of qualified points of view.

What of those who endorse a principle of public reason, but construe it differently? The reciprocity rationale for exclusion doesn't apply to advocates of the coercion model, since they accept public reason. Yet if we count them as fully qualified, the worry about self-defeat resurfaces. Since the reasons model applies to itself, inclusion of coercion-model advocates in the constituency of fully qualified perspectives would seem to make the reasons model fail its own test.[38] However, when we talk about a principle being rejectable, we always need to specify the alternatives, i.e. we need to answer the question "rejectable in comparison to what?" Suppose that we have a generic statement of the public justification principle P, which admits of two interpretations P_1 and P_2, e.g. the reasons and coercion models of public reason. P_1 is reasonably

rejectable *as against P₂*, but not as against the absence of any principle of public reason at all (not-P). The fact that it is reasonable to reject P_1 in favor of P_2 and P_2 in favor of P_1 is consistent with it being unreasonable to reject the disjunction [P_1 or P_2] as against not-P. Controversy about the specification of the public reason principle therefore doesn't make the principle self-defeating.

7.   Conclusion

I have argued that anarchy is only a danger for the coercion model of public reason, one that may be avoidable via an aggregation criterion that has a principled rationale; that self-defeat is only a danger for the reasons model, one that is avoidable by appeal to reciprocity; that idealization is in the first instance part of the problem to which public reason is a response, not a device for avoiding anarchy; and that for this reason among others the distinction between reasonable and unreasonable points of view does not involve any invidious differential treatment. The idea of public reason may be wrongheaded for other reasons, but it is not incoherent.

Endnotes

1   *Self-defeat*: Thomas Nagel, 'Moral Conflict and Political Legitimacy,' *Philosophy and Public Affairs*, 16, (1987), 222; Joseph Raz, 'Disagreement in Politics,' *American Journal of Jurisprudence*, 43, (1998), 30-32; Steven Wall, 'Is Public Justification Self-Defeating?,' *American Philosophical Quarterly*, 39, (2002): 385-94; *anarchy / minimal-state libertarianism*: Michael Sandel, *Liberalism and the Limits of Justice, 2nd Edition* (Cambridge: Cambridge University Press, 1998), 202-10; Simon Caney, 'Liberal Legitimacy, Reasonable Disagreement and Justice' in *Pluralism and Liberal Neutrality*, ed.

Richard Bellamy, and Martin Hollis (Illford, Essex: Frank Cass Publishers, 1999), 23-24; Joseph Chan, 'Legitimacy, Unanimity, and Perfectionism,' *Philosophy and Public Affairs,* 29, (2000), 20-22; *unequal treatment*: Philip L. Quinn, 'Political Liberalisms and Their Exclusions of the Religious' in *Religion and Contemporary Liberalism*, ed. Paul J. Weithman (Notre Dame, IN: University of Notre Dame Press, 1997), 138-61; Nicholas Wolterstorff, 'Why We Should Reject What Liberalism Tells Us About Speaking and Acting in Public for Religious Reasons' in *Religion and Contemporary Liberalism*, ed. Paul J. Weithman (Notre Dame, IN: University of Notre Dame Press, 1997), 162-81; Raz, "Disagreement in Politics"; Richard Bellamy, *Liberalism and Pluralism : Towards a Politics of Compromise* (London; New York: Routledge, 1999); Daniel M. Weinstock, 'Saving Democracy From Deliberation,' in *Canadian Political Philosophy: Contemporary Reflections*, ed. Ronald Beiner, and Wayne Norman (Oxford: Oxford University Press, 2001), 78-91; Erin Kelly, and Lionel McPherson, 'On Tolerating the Unreasonable,' *The Journal of Political Philosophy*, 9, (2001), 38-55; Michael Perry, *Under God? Religious Faith and Liberal Democracy* (Cambridge: Cambridge University Press, 2003).

2   David Enoch, 'The Disorder of Public Reason,' *Ethics,* 124, (2013), 141-76; David Enoch, 'Against Public Reason' in *Oxford Studies in Political Philosophy*, ed. David Sobel, Peter Vallentyne, and Steven Wall (Oxford University Press, 2015), 112-44. I will use the terms "public reason" and "public justification" interchangeably. All references to "Against Public Reason" will be made parenthetically in the text.

3   David M. Estlund, *Democratic Authority: A Philosophical Framework* (Princeton, NJ: Princeton University Press, 2008), 3-5, 40-64.

4   David Hume, *Essays Moral, Political, and Literary*, vol. 3, The Philosophical Works of David Hume (Edinburgh: Adam Black and William and Charles Tait, 1826), 519-20.

5   Enoch, "The Disorder of Public Reason," 170-3.

6   To be fair, advocates of what I am calling the coercion model have generally espoused weaker, "moderate" standards of idealization." See, for example, Kevin Vallier, *Liberal Politics and Public Faith: Beyond Separation* (New York: Routledge, 2014), 145-80.

7   This section of the paper draws on Andrew Lister, 'Public Justification of What? Coercion Vs. Decision as Competing Frames for the Basic Principle of Justificatory Liberalism,' Public Affairs Quarterly 25, no. 4 (2011): 349-67, and Andrew Lister, *Public Reason and Political Community* (London: Bloomsbury, 2013), 15-23, 81-105.

8   There is a question about whether all state action counts as coercive. If not, and the principle only applies to coercion, then the principle would not apply to all state action. Alternatively, one could frame the principle as a constraint on all state action, coercive or not; c.f. Colin Bird, 'Coercion and Public Justification,' *Politics, Philosophy & Economics,* 13, (2014): 189-214.

9   See in particular Gerald F. Gaus, and Kevin Vallier, 'The Roles of Religious Conviction in a Publicly Justified Polity: The Implications of Convergence, Asymmetry and Political Institutions,' *Philosophy and Social Criticism,* 35, (2009): 51-76; Gerald F. Gaus, 'Coercion, Ownership, and the Redistributive State: Justificatory Liberalism's Classical Tilt,' *Social Philosophy and Policy,* 27, (2010): 233-75; Vallier, *Liberal Politics and Public Faith: Beyond Separation*, 103-44; Kevin Vallier, 'In Defence of Intelligible Reasons in Public Justification,' *The Philosophical Quarterly,* (2015). The account provided in *The Order of Public Reason* is more complicated, because in this work Gaus aims to show that public justifiability is necessary to the authority of ordinary social morality, before turning to its role in relation to political authority and the state; Gerald F. Gaus, *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World* (Cambridge:

Cambridge University Press, 2010). Vallier's view includes a reasons constraint, since he requires that reasons be "intelligible." However, intelligibility is meant to be a relatively weak constraint, one that "expands the set of justificatory reasons to its outer limit;" Vallier, "In Defence of Intelligible Reasons in Public Justification," 7.

10 The phrase 'total balance of reasons' refers to the balance of reasons from a particular doctrine's point of view, including both the reasons that are distinctive to the view in question, and the reasons all qualified points of view accept (not counting the principle of public justification itself).

11 The distinction between the two models is closely related to the distinction between consensus and convergence justification. However, consensus is usually understood to require agreement on action based on agreement about reasons, which involves a double unanimity requirement; see, for example Kevin Vallier, 'Convergence and Consensus in Public Reason,' *Public Affairs Quarterly*, 25, (2011), 263. In contrast, the reasons model only requires agreement on reasons, allowing for disagreement about what laws and policies they support, or indeed about whether on the issue at hand it is preferable to have a common policy at all.

12 Enoch, "The Disorder of Public Reason," 170.

13 There is some disagreement on terminology in the literature on the self-defeat objection. David Estlund refers to a qualified acceptability requirement that meets its own standard of acceptability as "insular." Raz and Wall call non-insular principles self-defeating. A proposition is self-defeating, in one sense, when its truth implies its falsity. The truth of a non-insular acceptability requirement does not imply that it is false, but only that it cannot be used in political justification. For this reason, Estlund prefers to call non-insular principles 'self-excluding'; Estlund, *Democratic Authority*, 54. Unfortunately, 'self-

excluding' is the term Raz uses to refer to an acceptability requirement framed so as not to apply to itself (Raz, "Disagreement in Politics," 31), whereas Estlund calls these nonreflexive principles 'dogmatic,' 57. Nothing hangs on these differences, since a principle meant to guide political decision-making that excludes itself from deliberation is practically self-defeating, even if not false-because-true.

14 Gaus, *The Order of Public Reason*, 227-28.

15 Enoch, "The Disorder of Public Reason," 173.

16 Enoch, "The Disorder of Public Reason," 159-60.

17 The term "nested inconclusiveness" is from Gerald F. Gaus, *Justificatory Liberalism: An Essay on Epistemology and Political Theory* (New York: Oxford University Press, 1996), 180-82. The term "higher-order unanimity" is from Chan, "Legitimacy, Unanimity, and Perfectionism," 22.

18 Gaus, "Coercion, Ownership, and the Redistributive State," 250; Gaus, *The Order of Public Reason*, 323. The optimality criterion rules out Pareto-dominated options (ones everyone agrees are worse than another eligible option), whereas the eligibility criterion rules out options some member of the public thinks morally worse than the default of not having any law, on the matter at hand.

19 Enoch, "The Disorder of Public Reason," 153-54.

22 David M. Estlund, 'The Insularity of the Reasonable: Why Political Liberalism Must Admit the Truth,' *Ethics*, 108, (1998), 254-55; Estlund, *Democratic Authority*, 53, 57-8.

23 Wall, "Is Public Justification Self-Defeating?," 390.

24 David Estlund and Jonathan Quong take this route; Estlund, *Democratic Authority*, 61; Jonathan Quong, *Liberalism Without Perfection* (Oxford: Oxford University Press, 2011), 181.

25 According to Rawls, "the role of the criterion of reciprocity as expressed in public reason...
    is to specify the regime as one of civic friendship," John Rawls, *Political Liberalism* (New
    York: Columbia University Press, 1996), li; see also xlv, 213, and 217. For other discussions
    of public reason and political liberalism that emphasize the relational aspect of the view's
    justification, see Blain Neufeld, 'Civic Respect, Political Liberalism and Non-Liberal
    Societies,' *Politics, Philosophy & Economics,* 4, (2005): 275-99; Kyla Ebels-Duggan, 'The
    Beginning of Community: Politics in the Face of Disagreement,' *Philosophical Quarterly,*
    60, (2010): 50-71; RJ Leland and Han van Wietmarschen, 'Political Liberalism and Political
    Community', *Journal of Moral Philosophy* (2106): 1-26. The argument based on respect for
    persons due to Charles Larmore and James Boettcher can also be interpreted in this fashion,
    if one emphasizes the importance of relations of *mutual* respect; see Charles Larmore, 'The
    Moral Basis of Political Liberalism,' *The Journal of Philosophy,* 96, (1999): 599-625; James
    Boettcher, 'The Moral Status of Public Reason,' *The Journal of Political Philosophy,* 20,
    (2012): 156-77.

27 The main idea of relational egalitarianism is that what is fundamental to social justice are
    egalitarian relationships; equal distributions of economic goods are important only to the
    extent they contribute to or help constitute such relationships. As Elizabeth Anderson says,
    the positive aim of egalitarian justice is "to create a community in which people stand in
    relations of equality to others," Elizabeth Anderson, 'What is the Point of Equality?,' *Ethics,*
    109, (1999), 289. Relational egalitarians need not be committed to the view that equality
    only matters where one has opted into a special relationship. It could be that duties of
    distributive justice apply wherever people are living in proximity, sharing space and
    resources, and thus inevitably interacting, in one way or another. Without some assurance of
    compliance on the part of others, however, some of these duties will not be binding; see

Andrew Lister, "Reciprocity, Relationships, and Distributive Justice," in *Social Theory and Practice*, 39 (2013): 70-94.

28 Making public reason conditional on reciprocity does not reduce the principle to the basis of a modus vivendi. A modus vivendi is an agreement that is unstable with respect to changes in the balance of power between the parties involved. A commitment to public reason might be stable in this sense but nonetheless conditional on reciprocity. I might be willing to abide by public reason even if I get to be much more powerful than you, but only if I have reason to think that you would be likewise committed (or if I have no reason to suspect that you would not).

29 For more on this subject, see Andrew Lister, 'Public Reason and Reciprocity' in *The Journal of Political Philosophy* (2017).

31 Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), 272-73.

32 Raz, "Disagreement in Politics," 33.

33 Jonathan Quong, "The Rights of Unreasonable Citizens," in *The Journal of Political Philosophy*, 12 (2004), 314-335.

34 In Chapter 21 of *Leviathan*, Hobbes said that there was "no obligation on any man which ariseth not from some act of his own; for all men equally are by nature free…" Thomas Hobbes, *The English Works of Thomas Hobbes*, vol. III (London: John Bohn, 1839), 203.

35 Locke opens Chapter 8 of his Second Treatise by saying that men being by nature free, equal and independent, "no one can be put out of this estate, and subjected to political power of another, without his own consent," John Locke, *The Works of John Locke in Nine Volumes, 12th Ed*. (London: Rivington, 1824), 394.

36 For Hobbes, this right was encompassed within the Right of Nature. For Locke, it was part of each individual's executive authority over the law of nature, and was a duty as well as a

right, including the duty to help preserve mankind, at least where it did not put one's own life in danger.

37   I use the term "moral costs" because "costs" by itself would be misleading. The costs of adherence to public reason are not a setback to one's personal interests, but to one's moral doctrine.

38   Thanks to an anonymous referee for raising this point.