

Against Public Reason

David Enoch*

1. By Way of Introduction: Frustration

I don't know of any other philosophical discussion that is quite like this: Talk to (broadly speaking) Rawlsians, and you are likely to get the impression that some kind of political liberalism, or a public reason account, is the only game in town. The only questions worth discussing, it seems, are *within* this framework, rather than *about* it. Theorists who reject this framework are often ignored, and the feeling one gets is that they *just don't get it*. Talk to many others (non-Rawlsian political philosophers, as well as philosophers whose main area is not in political philosophy exactly), and you are likely to get the impression that Rawlsian public reason has been effectively refuted several times over, and indeed, that even this much was never necessary, as the theory was a non-starter to begin with. And the feeling one gets is that Rawlsians *just don't get it*, and that their tremendous influence in political philosophy is corrupting the field.

As my title suggests, I am much closer to the second stance than to the first (though both have been presented here in rather extreme, somewhat caricatured, forms). So in this paper I make yet another attempt to explain why public reason accounts are hopeless. I share the feeling that there are already in the literature strong reasons to reject specific public reason accounts and perhaps that tradition more generally, but I think that more can be done – and

* Earlier versions of this paper were presented at the ANU, Aarhus University Copenhagen, Leeds, Arizona, Yale, CUNY, Rutgers, Tel Aviv, and Jerusalem. I thank the participants on those occasions for the discussion and their valuable comments. And for extremely helpful written comments, I would like to thank Dani Attas, Dan Baras, Ilan Ben Shalom, Steve Darwall, David Estlund, Ruth Gavison, Alon Harel, Shelly Kagan, Barak Medina, Shmulik Nili, Shai Perry, Jon Quong, Chad van Schoelant, Steve Wall, Alex Zakaras, and an anonymous reader for *Oxford Studies in Political Philosophy*.

that given the state of the field, more should be done. In this paper, then, I try to develop some of the strongest, most general objections to public reason accounts¹. For this reason, I do not focus on the details of specific public reason accounts², not even Rawls's. Furthermore, in an attempt to get past the *they-just-don't-get-it* predicament, I try to broaden my vision, and to place some of the controversies here in a wider philosophical context. Also, I try to pinpoint – from the point of view of a public-reason-skeptic – the underlying intuition that public reason accounts *get right*, and to accommodate it without following them to their (unacceptable, I argue) conclusions.

In my more optimistic moments, then, I hope that this paper not only presents sound arguments, but also that it will advance the debate. But I am not unrealistic: I fully expect – in the descriptive, not the normative, sense – to come up against *he-just-doesn't-get-it* responses.

At the end of the next section, after presenting what I take to be the main underlying intuitions that push people in the direction of public reason accounts, I present a general characterization of the public reason accounts that will be my target in the rest of the paper. Public reason accounts, as I characterize them, are the set of views that are best seen as motivated by these or closely related intuitions, and that as a result endorse some requirement to justify political principles³ to each of those subject to them as a necessary condition for legitimacy. I then proceed, in section 3, to note that all public reason accounts must involve some idealization – because there is pretty much nothing citizens in modern societies all agree on, the requirement to justify political action to all requires (if anarchism is to be avoided) either

¹ Two I don't address here: self-defeat (which I address, following others to an extent, elsewhere; see Wall (2002), and my (2013)); and a critical discussion of the epistemic commitments of public reason theorists, which I intend to address in future work.

² I do some of this elsewhere, with regard to Estlund (2008; see my 2009), Gaus (2011; see my 2013), and to an extent Waldron (1999; see my 2007) (In what follows, from time to time I borrow a few sentences from these other texts of mine). Almost all other critics of this tradition target primarily Rawls, of course.

³ For my purposes here it won't be necessary to distinguish between public reason theorists who (following Rawls) restrict the scope of this requirement to just, roughly speaking, constitutional essentials, and those who (like Quong) apply the requirement more broadly.

restricting the scope of those to whom justification is owed, or idealizing the conditions under which such justification is owed, or both. General thoughts about when such idealization is a legitimate theoretical move and when it's objectionably ad hoc – as well as reflection about the specific idealization involved in typical public reason accounts – show that both options are subject to devastating critiques. In section 4 I defuse another one of the motivating thoughts of public reason accounts by discussing what reason we act on when we impose on others a principle we (but not they) believe in. I insist that the relevant reason for action is *the content of the principle*, not *that we accept it*. This means that the impression sometimes highlighted by public reason theorists that by flouting a public justifiability requirement we unacceptably give our own views extra weight is confused. In section 5, I take a step back from the details of some of the earlier sections, reflecting on the role of the political philosopher according to public reason theorists. I highlight what I take to be a deep, dangerous mistake – among other things, this conception of the role of political philosophers divorces them entirely from even the just struggles of political activists. In the concluding section I return to the starting point – to the powerful underlying intuition that public reason theorists, I think, get right. I show how such intuitions can be accommodated without endorsing a public reason account.

2. The Underlying Public reason Intuitions

Suppose that Catholicism is the true religion, and that the Pope has a direct line to God. Still, if you are a non-believer (or a believer of some other religion), we find it deeply objectionable for the Pope's directives to have authority over you, *simply in virtue of* the truth of Catholicism and the Pope's excellent epistemic credentials⁴. Certainly, the thought that you can be politically subject to the Pope's directives – backed up by the state's coercive power – seems

⁴ Estlund's (2008, 5) example.

unacceptable. The reason, it seems, is that while his directives are in some sense justified, they are not justified *to you*, non-Catholic as you are. And in order to reconcile legitimate authority with your nature as a free, autonomous agent, there must be something to be said for it, and furthermore, this something must be sayable *to you*. Truth, it seems, is never sufficient for legitimacy. Rather, those over whom authority is claimed are entitled to answers, to justifications of the authority that are available to them, pretty much as they are.

The general thought seems to be that given the liberty, autonomy, rationality, or some such of those subject to the purported authority, its legitimacy must be somehow *accessible* to them, it must be such that they too can appreciate it, it must be such that they have reason to endorse it, or, as the slogan goes, it must be justifiable *to* each and every one of them. It is impossible, the thought seems to be, for a free person to be subject to a legitimate authority if there's nothing that can be said *to him*, in terms *he* can relate to, as it were, that justifies subjecting him to that authority. Notice that no consent is required – perhaps, for instance, it's not about engaging your *will* (as consent presumably is) but your *reason* (a central distinction in Waldron (1987).)⁵ Still, the natural thought is that unless an authority can be justified *to you* pretty much as you are, it does not have legitimacy over you.

For brevity, I am going to put this underlying thought as the attempt to reconcile authority with liberty. Of course, these terms (“authority” and “liberty”) are neither univocal nor philosophically transparent⁶. But for my purposes here the somewhat loose characterization above suffices – clearly, something like the tension between authority and liberty does serious work here, as otherwise it would be entirely mysterious why we should justify an authority *to*

⁵ In fact, we can think about consent as a particular instance of this more general, arguably more accurate, condition. For if you've consented to an authority, we are guaranteed to have something to say *to you* about its legitimacy: Namely, that you've given your consent. According to this line of thought, then, the problem with consent theories is simply that they over-generalize.

⁶ For my take on authority in general, see my “Authority and Reason-Giving” (forthcoming).

those subject to it rather than to others. It is they who are owed justification, because it is their liberty (or some such) that is at stake when they are subjected to an authority. In my criticism of public reason below, nowhere will I rely on a more specific, more controversial characterization of the notion of liberty⁷ or of the need to reconcile liberty and authority.

Furthermore – and this brings us to the second underlying public reason intuition – for the Pope (or any of his followers) to apply his catholic-based directives to you non-believer as you are, will amount to giving more political weight to his (or their) beliefs than to yours, and so will amount not only to failing to treat you as free, but also to failing to treat you as an *equal* citizen, one whose beliefs, principles, desires, projects are as politically significant as anyone else's. When we combine these two underlying intuitions, we get the slogan – the justifiability-to requirement follows from the commitment *to treat each other as free and equal*.

This, then, is what in this paper I take to be characteristic of public reason theories. They are all committed to some kind of accessibility requirement as a necessary condition for legitimacy: The thought that for a state (for instance) to be legitimate, its authority must be justifiable to each of those subject to it. And – and this will be important in what follows – they all share the underlying motivation of reconciling liberty and equality with authority. To an extent, you should feel free to treat this characterization as a stipulation – in what follows, I will use the words “public reason accounts” to designate theories that are committed to some justification-to condition, motivated by the need to reconcile authority with liberty and equality. But I don't think that this way of using words is at all idiosyncratic, and I think that this characterization nicely captures pretty much all and only accounts that are usually thought of as members of this tradition. Here, for instance, is Nagel (1991, 33-6)⁸:

⁷ In particular, I will not be assuming a general freedom-to-do-as-one-chooses conception of liberty.

⁸ Quoted in a similar context in Rossi (2014).

The task of discovering the conditions of legitimacy is traditionally conceived as that of finding a way to justify a political system to everyone who is required to live under it ... the search for legitimacy can be thought of as an attempt to realise some of the values of voluntary participation in a system of institutions that is unavoidably compulsory.

And similar characterizations of the public reason account and underlying motivations are common⁹.

Now, as noted above, my hope is to criticize the entire public reason tradition, not any specific members thereof. So I cannot rely on more specific details – say, the exact nature of the accessibility or justification-to requirement¹⁰. Though such differences may be important in other contexts, for my purposes here what's important is what such different accounts have in common. In particular, then, I can afford not to say much more about what exactly it is that is needed for it to be the case that something is not just justified, but is justified *to* the relevant constituency. Still, one thing must be emphasized here: Though I am going to remain neutral on the precise nature of that extra thing that is needed (for justification-to), I am going to assume that this requirement is non-vacuous – that is, that a justification that is sound in an important

⁹ For statements of an accessibility or justification-to requirement as the definitive feature of this tradition, see, for instance, Waldron (1987, 128); Quong (2011, 161); Quong (2013, the opening line); Gaus's (2011, 263) "Basic Principle of Public Justification"; Vallier and D'Agostino's (2013) Public justification Principle; and Wall (2002, 385, 387).

For a helpful survey of possible motivations for public reason accounts, see Quong (2013, section 1), and the references there. There, thoughts about coercion constitute only one section out of five. But it is, I believe, the central one, and much of what goes on in the other sections also tacitly relies on it. For instance, another motivation Quong brings from Gaus has to do with Gaus's account of our reactive attitudes, and some accessibility being necessary for their appropriateness. But for Gaus, the appropriateness of the reactive attitudes is closely tied to thoughts about when the relevant part of morality is objectionably authoritarian – that is, to an extent, inconsistent with the liberty and autonomy of those subject to it.

¹⁰ For a helpful list of relevant distinguishing questions here (the nature of the accessibility, the constituency, the nature of the thing to be justified, etc.) and for many references, see Vallier and D'Agostino, Section 2.

sense can be accessible (in the relevant sense) to some but not to others, that the justification-to requirement takes as input something about the actual features of the relevant addressee¹¹. For now, this will do¹².

3. Idealization

I start my discussion of the role of idealization in public reason accounts by arguing that all (non-anarchist) public reason accounts need to employ some idealization (3.1). I then have an interlude on when in general idealization is a legitimate philosophical move (3.2). With a general answer to this question at hand, I proceed to reject the two idealization devices employed by public reason accounts – excluding the unreasonable (3.3) and going hypothetical (3.4). Similar objections, I conjecture, will apply to any other idealization device employed in order to save public reason accounts¹³.

3.1 Why No Public Reason Account Can Do without It

¹¹ Raz (1998) claims that the only accessibility requirement that makes sense here is one that is always trivially met, because all good normative arguments are always accessible to all. I don't agree with Raz on this – I think that there is sense to be made of the thought that the Pope's justifications are not accessible to a non-believer, even under the assumption that Catholicism is true. But my point in the text is just that an account that superficially uses the justification-to lingo, but which happily takes on board this point from Raz, is not within the target of my arguments in this paper (nor is it naturally classified as a public reason account).

¹² The words "public reason" are sometimes used in a different – narrower, perhaps even more precise – sense, one having to do with the shared language of officials and the Rawlsian duty of civility not to rely on private conceptions of justice. See Quong (2011, 41-2), and the references there. And for the distinction between the two ways of using "public reason", see Quong (2011, 256). I have nothing here to say about this other sense of "public reason" – I just note it here in order to preempt confusion.

¹³ Indeed, arguably the nature of the idealizing device doesn't much matter for the points to come. In particular, there is structural similarity between excluding the unreasonable and going hypothetical, perhaps to the point that a more theoretically parsimonious public reason theory will make do with just one of those (say, packing enough into the notion of the "reasonable" so that once the constituency has been restricted to the reasonable, no further hypotheticalization is needed). I discuss these two mechanisms mostly because public reason theorists employ both, usually in tandem. It should come as no surprise that similar worries arise regarding both.

Recall the characteristic feature of public reason accounts – they require that the relevant authority or principles be justified *to all* those subject to the authority. But this creates a problem, at least in the context of hoping to vindicate some contemporary states. The problem is that actual citizens of actual large-scale contemporary states are a very varied bunch. Different people are committed – sometimes even in the deepest ways – to all sorts of views and doctrines, they value – even intrinsically – all sorts of different things. If the justifications offered to them are to engage them as they actually are – perhaps based on principles they accept, or on the values they hold dear, or on what is already there in their motivational set – then it’s hard to believe that there is *anything at all* that can be justified to all. This is perhaps clearest on consensus-versions of public reason accounts, according to which for a political principle (e.g.) to be legitimate there must be a justification for it that is available (in the relevant way) to all¹⁴. But it remains true even on convergence views, according to which the condition necessary for legitimacy is just that for any citizen, there’s a justification available to her (without the further requirement that it must be the *very same* justification that’s available to all)¹⁵. So long as the justification-to requirement is non-vacuous, and so long as the relevant constituency consist of all the citizens of a contemporary state as we actually find them, it’s hard to imagine anything at all passing the bar. Anarchism follows.

Since public reason theorists are not anarchists, they idealize, thereby making the justification-to requirement easier to meet. The two ways of doing so that are used by public reason theorists are, first, restricting the scope of the relevant constituency (those to whom public justification is owed as a necessary condition for legitimacy) – so that what’s needed for

¹⁴ See, for instance, Larmore (2003), Quong (2011).

¹⁵ See, for instance, Gaus (2011).

legitimacy is not justifiability to all, but, say, to all the reasonable, or all the qualified¹⁶; and second, going hypothetical, so that those engaged by the justification-to requirement are not people as they actually are, but some hypothetical, idealized version thereof – perhaps, for instance, what’s important is not that people see the justification of the relevant principle, but that they *would* see or endorse it, *if* they spent a respectable amount of good reasoning on it¹⁷.

3.2 Interlude: When is Idealization an Acceptable Philosophical Move?

A theory is offered, one that ties some phenomenon to our relevant responses. Perhaps, for instance, a theory is offered about the relations between (phenomenal) color and our color appearances, so that to be red is just to appear red. Or perhaps a theory of values is offered that ties them very closely to what we value, so that to be of value just is to be valued. But counterexamples immediately come up: Sometime something appears red to me even though it isn’t, and sometimes it doesn’t even though it is. Sometimes people value things that aren’t of value, and fail to value things that are. A natural move then is to idealize: Perhaps to be red is not to appear red, but rather to appear red *to normal observers, in good lighting conditions*. Or perhaps to be of value isn’t to be valued, but to be valued *by the right people, in the right conditions*.

There is something perplexing about such a move. If you are really thirsty, but can’t find any water, offering you some ideal or hypothetical water (water that is not available, but would

¹⁶ Rawls, and following him pretty much everyone else, talks of reasonableness. Talk of qualified acceptance comes from Estlund (2008).

¹⁷ See Gaus (2011, 250).

have been available, if ...) will not be a good substitute¹⁸. Why are hypothetical or idealized responses any better?

The answer is that whether they are depends on what the underlying motivation was for going for the relevant view in the first place. Think again about the water example. Presumably, your reason for looking for water is that it would quench your thirst. Hypothetical water won't do that – it won't even go a part of the way towards doing that. So in the absence of water, going for hypothetical water amounts to cheating – it is disconnected from the underlying motivations of looking for water, and therefore offers no satisfaction, not even partially.

What about colors? Is idealizing here a way of cheating, avoiding counterexamples in an ad hoc way? Or are ideal, hypothetical responses enough here? The answer depends on the philosophical motivations underlying the relevant account of colors. If they are all about actual observers and their actual responses, then going hypothetical (because of the pressure from counterexamples) is cheating. What counterexamples show us, in such a case, is not that we should settle for observations in hypothetical conditions, but that we should resist the attempt to tie colors and our appearances of colors as closely together as all that. But if the underlying motivations of such an account are consistent with settling for hypothetical conditions, then all is well, and the idealized response-dependence account may still be a good idea.

Similarly for values and what we value. If the underlying motivations for offering a response-dependence view of values is tied to actual people and their responses, then what we should do in the face of the obvious counterexamples is not idealize (in an ad hoc way), but rather reject response-dependence altogether. But if the idealization can be motivated in a way that's consistent with going response-dependence and with the philosophical motivations for so going, then all may still be well.

¹⁸ The line here echoes a well-known criticism of hypothetical consent theory, for instance, in our context, Dworkin (1973).

And so, we have a test for when idealization is a legitimate philosophical move. It is, when it is motivated, and furthermore, when the offered motivation is consistent with the motivations for going for the initial, non-idealized view (the one that was devastated by obvious counterexamples). This is why hypothetical water is out, why (perhaps) some idealized response-dependence views of colors may be in, and why (as I argue elsewhere) idealized response-dependence views of normative concepts are out¹⁹.

Now we can put the test to use with regard to the idealization that is part and parcel of all public reason accounts. Is it legitimate, motivated, idealization? Is it consistent with the underlying public reason motivations? Or is it ad hoc, the kind of idealization whose sole purpose is to artificially save a theory from obvious counterexamples?

3.3 Excluding the Unreasonable²⁰

Think, then, about the restriction of the justifiability-to requirement to just the reasonable (or some such), so that the fact that a principle is not justifiable to the *unreasonable* does not undermine its legitimacy. What is the motivation for this restriction? Can it be reconciled with the underlying motivation for public reason accounts?

Though what is meant by “reasonable” (even just by Rawls) is a matter of controversy I’d rather not enter, still I think we can say, for our purposes here, that public reason theorists typically think of reasonableness as consisting of a motivational and an epistemic element. The epistemic element is acknowledging things like “the circumstances of justice”, reasonable pluralism, or the burdens of judgment and their effects – according to such conditions, if you

¹⁹ See my “Why Idealize” (2005). But see also Sobel (2009) and my (manuscript).

²⁰ Worries in this vicinity have been around since very early on. See, for instance, Hampton (1993, 299), and Raz (1998, for instance 37). In what follows, I try to address them both more generally (not just vis-à-vis Rawls), and in a more principled way (based on the general thoughts in the text about when an idealization is philosophically acceptable).

just expect (in the descriptive sense) everyone to converge on your religion, say, you're not being reasonable. The motivational element is roughly about wanting to interact with others as free and equal, perhaps even being more strongly motivated by such a desire than by many others²¹. If you don't think of others as your equals for political purposes, if you're entirely happy interacting with them by manipulating them rather than by reaching agreements, then you're not reasonable.

Notice that "reasonableness" here is a *technical* term, introduced to serve a certain role in the theory by stipulation. Sure, it may bear some resemblance to how we use the term in natural language, but the characterizations of reasonableness (for instance, in the previous paragraph) are a matter of stipulation, not of conceptual analysis²². So – their many attempts to the contrary notwithstanding²³ – those invoking the concept cannot rely on intuitions we supposedly have about who is and who isn't reasonable. We don't have such intuitions about technical terms. This is not just a methodological glitch. For often, perhaps by relying on some pre-theoretical intuitions about a non-technical concept of reasonableness, public reason theorists write as if the unreasonable are pretty much the Nazis and the murderous psychopaths²⁴. Those, of course, are unreasonable also in the natural-language meaning of the term; and frankly, excluding *them* doesn't hurt that much, and perhaps – though I'm not as sure – is not theoretically problematic either. But this is cheating. While the Nazis and the murderous psychopaths are obviously unreasonable, so are – on public reason accounts – many, many others. For instance, at least some of the Rawlsian thoughts about the burdens of judgment and their epistemic significance would be rejected by pretty much any epistemologist working in the

²¹ See, for instance, Quong's (2011, 291) understanding of (Rawls's understanding of) the reasonable.

²² In this respect, Estlund's insistence on going for a more technical term ("qualified acceptability") – motivated precisely by the kind of considerations in the text here – is more helpful.

²³ For an especially clear example, see Quong (2011, 156).

²⁴ Quong's examples, for instance, are those making "claims to racial, gender, or ethnic superiority" (292), the Ku Klux Klan (299), white supremacists (309), Nazis (309), and psychopaths (314).

field today²⁵. Perhaps more importantly, if you are not a Rawlsian – if you don’t accept something like a public-reason justification-to requirement as a necessary condition for legitimacy – then this too suffices, according to public reason accounts – to make you unreasonable²⁶. And – perhaps depending on the details – this may be so even if you do go for a public reason account, just not the right one²⁷. So it’s really important to keep in mind – when discussing the status of the unreasonable – that we’re not talking about just the Nazis and the murderous psychopaths. For any public reason-theorist, the ones excluded as unreasonable – that is, the ones justification to whom is not necessary for legitimacy – arguably include also such people as John Stuart Mill, Karl Marx, Joseph Raz, Jean Hampton, pretty much all contemporary epistemologists, probably most of those offering a rival public reason account, presumably the early Rawls – oh, and me.²⁸

Now, a case can be made that these aspects of the delineation of the unreasonable are a necessary feature of public reason accounts²⁹. But, loyal to my attempt to target the tradition in general rather than some (even all developed) specific manifestations thereof, let me not commit myself to this strong claim. So the thing to keep in mind is not necessarily that the

²⁵ I hope to address the shaky epistemological-sounding claims made by Rawlsians on another occasion.

²⁶ For explicit claims to this effect, see Dreben (2003, 326); Estlund (2008, 61); Quong (235, fn 34; 240). As far as I know, Rawls nowhere says this explicitly. Nor do I know of a place where he rejects this claim.

²⁷ For instance, I think – but I’m not entirely sure – that according to Quong’s consensus view, a Gaus-like convergence view is unreasonable.

²⁸ Public reason theorists can insist (as Quong (2011, chapter 10) explicitly does) that the unreasonable should enjoy the same rights that the reasonable enjoy – it’s just that which rights these are is going to be determined by justifications that are accessible only to the reasonable. Still, even when just talking about the distinction between those to whom justification is owed and those to whom it is not owed it is important to keep in mind that the latter include many who are, in the natural-language meaning of the term, perfectly reasonable.

I should note that Quong’s commitment to equal rights to the unreasonable is not without qualification. Quong is willing to seriously restrict the freedom of speech of the unreasonable (309). I am not sure, but I think that Quong may be committed to the justifiability of preventing me from publishing this article, for instance.

²⁹ At least one plausible way of going on the problem of self-defeat has it come out as a theorem that all non-public-reason-theorists are unreasonable. But perhaps there are ways of avoiding such a result.

reasonable include all non-Rawlsians, but that they include many more than just the Nazis and the murderous psychopaths³⁰.

Can public reason theorists retreat to a pre-theoretical understanding of the reasonable, perhaps so that you qualify as reasonable if your reasoning mechanisms are functioning well, or if you're willing to listen to other views, and so on³¹? They can, of course, but the price would be anarchism again – for among those who are reasonable in just some very thin sense of this kind, *everything* is controversial. Under any plausible hypotheticalization, and any non-trivial understanding of accessibility, *nothing* is justifiable to all the reasonable in *this* sense.

Keeping in mind, then, how easy it is to be unreasonable, let's return to the philosophical motivation underlying public reason accounts – the need to reconcile authority with liberty and equality. And we must ask – do such thoughts apply to the unreasonable as well? The answer seems obvious. The unreasonable too are born free, and yet they are everywhere in chains³². And if by coercing someone based on a principle they do not endorse (nor would they, after a respectable amount of good reasoning) we fail to treat her as our equal, then this is so when that someone is unreasonable as well. If you're not sure, recall the need to resist the temptation to think of the unreasonable as the Nazis and the murderous psychopaths. Suppose I am quite, well, reasonable (in the non-technical, natural language sense of this term), except I reject some of the burdens of judgment. Is there no longer even an initial tension between my liberty (in whatever sense is important here) and the authority the state claims

³⁰ The point is not that there's something strongly counterintuitive about classifying these people as unreasonable. The point, rather, is first, that the public reason theorist can't have it both ways – if he applies every-day natural-language intuitions regarding "reasonable" in ways that support his theory (which he shouldn't do, as "reasonable" here is a technical term), he must say something about the cases in which they clearly challenge it; and second, that in the technical sense of "unreasonable" – namely, those justification to whom is not a necessary condition of legitimacy – excluding all those people is indeed counterintuitive.

³¹ Sometimes, public reason theorists write as if they do. See, for instance, Nagel (2003, 76); Larmore (2003, throughout).

³² I'm paraphrasing Rousseau here, without any exegetical pretensions.

over me? Suppose I am motivated to interact with others as free and equal, but I have other stronger motivations, or perhaps my conception of what it is to treat others as free and equal is a little bit off. Is there no longer a problem of equality if you impose on me a principle I do not endorse? Alternatively, just think of some of the smartest, nicest, people whom public reason theorists rule out as unreasonable, and ask whether the underlying thoughts about liberty and equality apply to them, at least initially. I can't see how a negative answer can even get off the ground. A story reconciling authority with liberty and equality is still very much needed, with regard to (at least many of) the unreasonable as with regard to everyone else.

If the initial problem arises with regard to the unreasonable as well, and if the restriction to the reasonable is not to be an ad hoc restriction introduced just in order to save the theory, some other rationale must be supplied for the exclusion of the unreasonable. For the most part, this is a need left unacknowledged by public reason theorists. Still, at times such rationales are offered (or can be read off what public reason theorists say). Let me address the two main ones I find in the literature³³.

It is sometimes said that reciprocity considerations are what does the trick here. The thought seems to be that we should justify ourselves only to those who, had they had political power, would have bothered to justify themselves to us³⁴. Now, it seems to me an interesting, general question how such reciprocity considerations in general behave, and I can't discuss this question in general here (nor do I know what I would say). Let me just quickly note two points. The first is that reciprocity considerations never seem to be at the most fundamental moral

³³ It is sometimes hinted that practical considerations – having to do, roughly, with the costs of including the unreasonable – do the work here. See, for instance, Cohen (2009, 7), Quong (2011, 37). But I don't think that these suggestions are in line with the public reason spirit – on top of relying on questionable empirical assumptions (seeing how many people are *unreasonable* in the relevant sense, excluding the unreasonable may also be extremely costly), it seems too *modus-vivendi-ish* for Rawlsians. This is not what they seem to be after.

³⁴ Macedo (1995, 484), Larmore (2003), Lenman (2010, 179).

level. Perhaps, that is, I should only give you a ride to work if you, had you had a car, would have offered me a ride. But when it comes to the most fundamental level, this is not so: If utilitarianism is true, then the persons whose utility we should take into account are all, not just those who would take our utility into account; and I don't recall, in Kant, the proviso that what we should always treat as an end and not merely as a means is the humanity of only those who would treat us in a similar way. And we are here in a most fundamental context indeed – the basis of all political philosophy, the need to legitimize state authority. Here too, then, reciprocity considerations just seem inappropriate. Those who would “misbehave” politically are still free and equal, and so subjecting them to authority still calls for justification. Second – to anticipate the next point – because so many of the constituency of even current Western democracies are unreasonable, reading the whole project as based on such reciprocity considerations threatens to divorce it entirely from the real-world.

According to another possible rationale for excluding the unreasonable the justification for the exclusion of the unreasonable is *methodological*. We are doing, at this stage, ideal theory, by which is meant (perhaps roughly) a theory that assumes full compliance. And it is natural and methodologically advisable to start off by assuming away some further complications³⁵. The point is underdeveloped in Rawls, but perhaps we can fill in some details: Even assuming a political community where everyone is reasonable (in the relevant sense), still the burdens of judgments apply; so still, reasonable pluralism is to be expected; and so we can ask – how are we to proceed, in *that* society, given such pluralism? The requirement to justify political principles (or some such) to each of the reasonable is the answer to *that* question. The

³⁵ See Rawls's “The Domain of the Political and Overlapping Consensus” (1989, 236). That he's doing ideal theory, and that ideal theory is prior to non-ideal theory, are of course central themes in Rawls's *A Theory of Justice*. But I don't know of claims in *Political Liberalism* where Rawls justifies excluding the unreasonable based on these general points.

unreasonable are irrelevant, because they are not a part of that society, or an addressee of our ideal theory.

Now, there's something that seems to me importantly right about this way of proceeding: Namely, the problem is still interesting even when the unreasonable are assumed away; and philosophers need no further reasons to engage a philosophically interesting question. Perhaps – though I'm much less sure of that – it's even a good idea to accord ideal theory methodological priority (in some sense) over non-ideal theory. But even granting Rawls all that, still we shouldn't be tempted by too quick a move from ideal to non-ideal theory. Even if Rawls is right that some kind of justifiability-to is a necessary condition for legitimacy in an all-reasonable-community, it most certainly does not follow that in a community where some are unreasonable, the analogous condition necessary for legitimacy is justifiability-to-just-the-reasonable (nor does it follow, if you've justified something to all the reasonable under full compliance, that it's also justifiable – even just to them – under partial compliance). Unless some rationale can be offered for this restriction *in the non-ideal theory*, then perhaps the thing to conclude is that no justifiability-to requirement is necessary for legitimacy in non-ideal theory, because none is satisfiable (because of the presence of the unreasonable). Or perhaps there are some other ways to go. The general point is simply that it's *never* a simple, immediate step from ideal to non-ideal theory³⁶. Ignoring this would be tantamount to trying to design airplanes by relying on mechanical models that assume frictionless surfaces, just because there's something to be said for thinking of some such models for some purposes. So, despite the intellectual respectability of the ideal-theory way of motivating conducting a discussion that

³⁶ For similar points in similar contexts, See Cuneo (2013, 360), and Rossi (2014). This seems to be a point Quong understands well, at least when criticizing others: "... even if we stipulate that people can only be members of the justificatory constituency if they accept Gaus's views on rational justification, this means Gaus's approach (and by implication the convergence model) does not apply to the world that we live in, nor does it apply to a well-ordered liberal society." (272-3)

ignores the unreasonable, this way of proceeding cannot at the end of the day vindicate excluding the unreasonable, at least when we're back to talking (as public reason theorists seem to) about the real world.³⁷

Notice that my point here – perhaps unlike some other current discussions of ideal and non-ideal theory in political philosophy³⁸ - strikes at the very heart of public reason accounts: their suggested necessary condition for legitimacy. The point is not just about the unsoundness of drawing more specific practical implications from an ideal theory to the real world (though it is that too); nor is my point about the interest in doing ideal theory (of course that too is interesting and valuable. Many things are.); nor is the point about the importance of aiming at the truth regarding political matters, even when doing so will have no effect in the world (I agree that this too is of importance, and a part of the political philosopher's job). The point, rather, is that if the only way for the public reason theorist to motivate excluding the unreasonable is by resorting to ideal theory, she can no longer maintain that justifiability to all the reasonable (but not to the unreasonable) is a necessary condition for legitimacy in the real world, where not everyone is reasonable. All she has to say, then, is that justifiability to all would be a necessary condition for legitimacy in an all-reasonable environment. I doubt that

³⁷ Dreben (2003) is a clear example of someone emphasizing both the ideal-theory nature of the project (e.g. 323, 344) and the thought that important conclusions about the real world follow (e.g. 317, 328). He doesn't do anything to reconcile these arguably inconsistent claims.

I think that Quong's recent (2011) emphasis on the *internal* nature of the political liberal project is in line with the Rawlsian ideal-theory way of excluding the unreasonable (In conversation, Quong has confirmed this.), and at times (e.g. 143) he even talks explicitly in terms of ideal theory. Now, I think that Quong's insistence on the internal project cannot withstand criticism in general, but he develops this line in great detail, and offers explicit justification for doing it. I remain unconvinced, and I hope to say why in future work. Here I must settle for noting that as a response to the problem of excluding the unreasonable, Quong seems to fail for reasons similar to the ones in the text here. See especially 290, where he seems to be guilty of precisely the fallacious move from ideal to non-ideal theory highlighted in the text – having justified something under full compliance, proceeding to assume that it's been justified to the complying under partial compliance.

³⁸ See, for instance, Farrelly (2007), Stemplowska (2008), and Simmons (2010). Unlike Simmons, I am not invested in any interpretive claim about how best to understand these terms in Rawls. And notice also that the point I am about to make in the text stands even if everything Simmons says (including about the priority of ideal theory, in the sense he finds in Rawls) is true.

many public reason theorists will be happy with this result³⁹, and I want to insist that they should not. (I'll have more to say about how going down this road betrays the role of the political philosopher later on.)

The underlying problem of reconciling the state's authority with the freedom and equality of those subject to it arises with regard to the unreasonable just as it does with regard to the reasonable, and no rationale that I can think of or that I find in the literature can motivate the exclusion of the unreasonable from the constituency of the public reason requirement consistently with the underlying public reason motivations. So excluding the unreasonable in this way is objectionably ad hoc – it is a revision of a theory, introduced in a way that's inconsistent with its underlying motivations, just in order to save it from counterexamples. So the thing to do given the obvious counterexamples is not to exclude the unreasonable, but to reject a justifiability-to requirement.

3.4 Hypothetical Conditions

So much, then, for the first mechanism employed by public-reason theorists to avoid the problem that if we take people as we find them, nothing is justifiable to all of them (on any plausible understanding of justifiable-to). The second mechanism mentioned above is that of going hypothetical. What is needed for legitimacy, the thought is, is not necessarily that the relevant principles (or whatever) be justified to the satisfaction of members of the relevant constituency as we find them, but rather that they *would* be satisfied, if placed in the right hypothetical conditions. Such conditions may include things like being fully-informed, perhaps

³⁹ Quong seems to be officially okay with this result, but – as can be seen from the references in footnote 37 above – doesn't seem consistent in this regard.

satisfying some norms of procedural rationality, thinking about things in a calm moment, and so on.

But now we should ask of this idealization, or hypotheticalization, whether it passes the test from section 3.2 above. Is it well-motivated, in a way that is consistent with the motivations underlying public reason accounts in the first-place? Or is it an ad-hoc “fix”, introduced merely in order to avoid some obvious counterexamples and save the theory?

Think again of consent and related normative phenomena. Perhaps I owe you some money, because of a contract I signed. You come asking for your money, and I ask why I should give it to you (myself, I would rather use it in different ways). You note that I consented, or promised, or signed a contract, or some such. This seems, at least sometimes, as a good response. But now suppose there never was any contract. Still, you are not at a loss for words. To my question (why should I give you the money) you respond: “I know you didn’t give your consent. But you would have, if you spent a respectable amount of good reasoning on it.”, or utilizing some other hypotheticalization of this kind. This sounds like no answer at all. In this case, hypothetical consent is no more substitute to consent than hypothetical water is to water for the thirsty. Of course, you may want to indicate that I should have consented, that I had overwhelming reasons to. But if you do that, it’s no longer clear what work is being done by the hypothetical consent – you can explain why I have to pay directly by appeal to these underlying reasons. This is a common, well-known problem for hypothetical consent theories⁴⁰. And a similar initial worry seems warranted in the context of a justification-to principle as well. It’s not clear how the fact that a justification *would have* satisfied me had I been in some hypothetical conditions goes any way at all towards reconciling authority and my liberty, given that I *am not*

⁴⁰ Again, see Dworkin (1973). For a recent attempt to respond to it – though in the context of his specific, somewhat idiosyncratic, *normative* hypothetical consent theory – see Estlund’s (2008) discussion of normative consent.

actually satisfied with the justification offered. True, perhaps I *should* be so satisfied. But then, presumably there are reasons why I should be satisfied with the justification offered, and then all the work is being done by those reasons directly, not by their availability to me. The justification-to requirement has been left out of the normative picture⁴¹.

If this were so – if no hypothetical satisfaction with an offered justification ever did any normative work – then public reason theorists would have to choose whether to go for actual satisfaction (and embrace the anarchism that would follow) or to concede the failure of the justification-to requirement and with it of the public reason tradition in general. But I think that this would be too quick. For sometimes, even in discussing consent, hypothetical consent does matter. Perhaps, for instance, mild, superficial corrections of procedural irrationality are okay in this respect – perhaps, in other words, if you don't consent to a medical procedure because you're too drunk to listen to the explanations you yourself would have accepted unhesitatingly a couple of drinks ago, then for some purposes the situation is similar to one in which you did give your consent⁴². So hypothetical consent – and hypothetical justifiability – are not just like hypothetical water. Sometimes they make a normative difference.

Ok then, but *when*? Under what conditions does the fact that one – dissatisfied as one actually is with the offered justification – *would* be justified in some suitably specified hypothetical conditions make the needed normative difference? I don't have a general answer to this question (I hope to have more to say on in the future). And so the discussion here will not be complete. But it seems to me that even in the absence of such a complete discussion, we can return to the circumstances of political disagreement and argue that hypotheticalizing here is of no help at all.

⁴¹ "There may be no middle way between actual (including implied) agreement and rational justification." (Raz, 1990, 46).

⁴² This often-used kind of example comes, for instance, from Waldron (1987, 132).

The way to see this is to think about a real-life political disagreement. The public reason theorist claims the status of legitimate authority for some principle on which political action is to be grounded – say, one allowing for a rather extensively redistributive scheme of taxation; a (reasonable) citizen comes along, rejecting the justification offered for the principle or policy⁴³; furthermore, she challenges the legitimacy of the principle, because it has not been justified to her, to her satisfaction; the public reason theorist is undeterred: for legitimacy what is required, he reminds her, is not that the principle be justified to her satisfaction, but that it be *justifiable* to her satisfaction, so that she *would* be justified, in the suitably specified hypothetical conditions. Perhaps, he insists, she hasn't been paying enough attention to the arguments, and would have been convinced had she done better in this regard; or perhaps she's being in some other way procedurally irrational; perhaps she's not well-informed; perhaps, while satisfying the minimal reasonableness criteria, still she hasn't been taking others' point of view into account sufficiently impartially. Because this is so, the public reason theorist reassures himself, there is really no legitimacy problem after all. All is well.

How convincing is this line of thought, as applied to real-life political disagreements? Remember, we're not talking about the highly anxious or drunk patient, in the emergency room, in an especially vulnerable time in his life. We're talking about serious people taking part in serious discussions over long periods of time, in the public sphere or in the privacy of smaller social interactions, and anyway very deeply committed to their respectable views. And let's set aside for now other related worries, such as how civil it would be to *say* to the one rejecting the principle that he would have accepted it had he spent a respectable amount of good reasoning

⁴³ See, for instance, Gaus (2011, Chapter 24).

on it⁴⁴. The question we're interested in is whether such a response is in line with the underlying public reason motivations. If I am the person rejecting the suggested taxation scheme, I am in the Rousseauian predicament – supposedly free, but about to be coerced to take part in this policy which I reject. Is this tension in any way reconciled seeing that I would have accepted the principle in some hypothetical conditions that are sufficiently far so that I am not moved here and now, not even having thought about things quite seriously, having debated them with my friends, and so on?

I am inclined, of course, to say that this is no response at all. The tension between authority and liberty is fully present in these cases, even given this kind of hypothetical justification-to. Now, let me emphasize again that I am not claiming that hypothetical consent and justification-to never matter. What I claim is that the thought that they matter *here*, where they are supposed to matter most, in the context of political disagreement, resists belief. Also, note that all of this is supposed to be from the point of view of a public reason theorist, one who thinks that justification-to is extremely important. *My* point is not that it's extremely important to justify the taxation scheme to all those rejecting it. My point is just that *if* you think that justification-to is a necessary condition for legitimacy, because this is the only way to reconcile authority with liberty and equality, or to respect as equal and free your fellow citizens, *then* you should not be impressed with the response at the end of the previous paragraph. That response amounts to no progress in respecting your fellow citizens as free and equal. It should not impress you more, for instance, than the response "I understand that you are not convinced, but I've already stated my reasons for why it's *true* that this taxation scheme is justified."⁴⁵ And this

⁴⁴ For this point, put in terms of adding insult to injury, and directed at Gaus, see my (2013, section 6). Hampton (1993, 309) seems to anticipate something like this line. And Wolterstorff (2012, 74) makes precisely this point as well.

⁴⁵ A similar point comes up in the context of granting conscientious exemption from military service. Statman (2009) argues that in the Israeli context pacifist objectors are wrong about what follows from

response is, of course, not one that a public reason theorist can settle for without abandoning his commitment to public reason.

Furthermore, the public reason theorist who wants to use such idealization owes us also the details – what the idealization does and what it does not rule out. It's not going to be easy to fill in the details in an extensionally adequate way. And what's more – extensional adequacy is not enough. It's also necessary to offer a rationale for the specific way of idealizing, or the specific hypothetical conditions deemed relevant, a rationale that's consistent with the underlying public reason motivations. I don't see how this can possibly be done.

Going hypothetical, I conclude, is no more promising a strategy for the public reason theorist than excluding the unreasonable. Any such attempt will be objectionably ad hoc. The public reason theorist should either settle for actual justification-to as a necessary condition for legitimacy (which in all likelihood entails anarchism, even when restricted to just the reasonable), or else take back his commitment to public reason.

4. The Reasons We Act On (For Instance, When Coercing)

Recall the thought that by refusing to go sufficiently neutral – by relying on claims that are not justifiable to the entire relevant constituency – we fail to treat those to whom it is not justifiable as equal. The thought seems to be that if I am willing to impose Catholicism-based imperatives on you, non-Catholic as you are, I am giving extra political weight to my own beliefs over yours. I am treating in an asymmetric way the fact that I believe Catholicism and the fact that you

their own pacifist commitments, so that refusing to exempt them is no attack on their conscience. I (2009b) respond by noting the point in the text: The tension between freedom of conscience and the demands of the state is not in any way relieved by the fact (if it is a fact) that the objectors are wrong about what follows from their own deepest commitments.

believe its denial. According equal weight to your beliefs and mine would require going neutral here.

This thought, though, is based on a confusion about reasons for actions⁴⁶. When I impose (in good faith) Catholicism-based directives on you, my reason for action – that is, the feature of the circumstances that I take to be normatively relevant – is not *that I believe Catholic doctrine*. Rather, my reason is the content of that belief, namely *Catholic doctrine* itself. Similarly, if I, a comprehensive, “metaphysical” liberal impose directives that are based on the value of autonomy on people who do not value autonomy, my reason for action is not *that I value autonomy*. Had this been the reason, then given that they do not value autonomy, privileging my valuing over theirs would have violated some intuitive principle of equality. But my reason for action is different, and is not about me at all. Rather, my reason for action is *that autonomy makes people’s lives go better*, or some such. There is nothing indexical about this reason for action, and so no violation of equality at all.

The point is tricky, and so merits more elaboration. It’s tricky partly because it’s hard to distinguish, from the first-person perspective, between *that-p* and *that-I-believe-that-p*. But with the help of some counterfactuals, this can be done. We’re assuming, then, that in the actual world, you think that autonomy is of value. Let’s imagine a possible world in which you are mistaken about autonomy, thinking that it’s not of value, even though it is. Do you still want – in the actual world – to impose autonomy-based directives on people in that hypothetical world? A positive answer seems obvious. But in that world, you don’t believe in the value of autonomy. Still, in that world autonomy is of value. So what we can learn from the fact that when you think (in the actual world) about that possible world you still want to impose autonomy-based

⁴⁶ This confusion pervades, I believe, Rawls’s and Rawlsian texts. For one clear example, see *Political Liberalism* (1993, 129). Usually, though, the confusion is less explicit than that. Still, attributing this confusion to Rawlsians is needed in order to make sense of claims that equality is in any way relevant here.

directives is that your reason for imposing them is *that autonomy is of value* (which is true there too), not *that you believe that it is of value* (because on that world, you don't). Compare this to cases of, say, taste. Suppose that in the actual world you prefer vanilla ice cream over chocolate. Now imagine a possible world in which you prefer chocolate. And suppose that you – in the actual world – have to order ice cream for your counterpart in that possible world. What flavor do you order? Clearly, chocolate. What this shows is that in the actual world too, when you order vanilla ice cream, your reason for action is indexical, it *is* about you and your preferences, it's simply *that you prefer vanilla*. This is why it makes no sense to order vanilla ice cream for your counterpart in a world in which he no longer prefers vanilla. But when you act on a principle or something like the value of autonomy, this simply isn't so. Your reason for action in such cases is *that-p*, not *that-you-believe-that-p*. It's just not about you at all. So there is no sense in which you're giving extra weight to your beliefs over others' – you're giving *no* weight to your beliefs here. You're just giving weight to the value of autonomy, or the truth of Catholicism, or some such.

This point is in no way special to the political realm. It applies across a wide range of cases, in practical as well as theoretical reasoning, regarding justification and explanation alike. Suppose you believe that what explains the tide is the gravitational force of the moon. What, do you believe, does the explanatory (and causal) work here – *that the moon has this-and-that gravitational force*, or *that you believe that it does*? Clearly, the former. You don't believe that your beliefs causally regulate the tides, nor do you think they explain them. What has relevant explanatory force is the content of your beliefs, not the fact that you have them. We can run the counterfactual test again – surely, when you think (in the actual world) about a possible world in which you're wrong in your beliefs about the explanation of the tides, you don't (in the actual world) think that the explanation of the tides in that possible world is any different than the one

in the actual world. What does the explanatory work is entirely present there – it is the moon’s gravitational force, not your beliefs about it (which are absent there). The explanation, then, is not about you at all.

Similarly, in epistemology the question sometimes arises whether – in a case of disagreement with someone you take to be your epistemic peer on the relevant question – you should give extra weight to your own evaluation of the relevant evidence. Here too, though, it’s clear that in typical cases, the thing to which you accord epistemic weight is not *that you evaluate the evidence a certain way*, but rather *the evidence and its epistemic value*. You are not a part of the picture of what does the epistemic work. It’s just not about you⁴⁷.

And the same point applies to our reasons for action even when they are not themselves normative. If I believe that cutting the blue wire will defuse the bomb, you believe that cutting the red one will, and I proceed to cut the blue one, there is no sense in which I am giving extra weight to myself or my beliefs. My reason for action is not *that I believe that cutting the blue wire will defuse the bomb*. This is not the feature of the case to which I accord normative significance (as can be seen, for instance, by running the counterfactual test again). Rather, my reason for cutting the blue wire is *that doing so will defuse the bomb*. And *this* reason is not indexical at all. Once again, then, it’s not about me.⁴⁸

⁴⁷ This is one of my main points against the Equal Weight View in “Not Just a Truthometer” (2011b). For similar points in a wider epistemic context, see Schroeder (2008). Of course, not many things are uncontroversial in the peer disagreement literature. But the point in the text, it seems to me, should be. (Those I criticize in “Not just a Truthometer” do not *argue* against this point; rather, they seem to either *assume* its denial, or not to see its full significance.)

⁴⁸ You may be worried about mistakes: What if in fact autonomy is not of value? Surely, in that case my reason for action can’t be that autonomy is of value (for it isn’t), and the only remaining candidate seems to be my belief in the value of autonomy. But this is not so. A fuller discussion of these issues – like the one I offer in (2011a, 221-3; see also the references there) – will distinguish between different senses of “reason for action”, only some of which factive. In the terms I use there, *the agent’s reason* – the feature of the situation the agent finds normatively relevant – need not be factive. Even in the case in which autonomy is not of value, then, my reason for action may be the value of autonomy (just like even if the moon is not after all responsible for the tides, still what in my view does the explanatory work is the moon’s gravitational force, not my belief about it).

Now, you may think that still there is a sense in which I am not out of the picture when it's the content of *my* beliefs that plays the relevant role in my reasons for belief or action. And I think that this is right, but we have to be careful about the precise nature of this role⁴⁹. The point is, again, that the fact that it is my view is not a part of what is, on my view, normatively relevant (unlike, for instance, in the mere preference for vanilla case). It is thus not a part of my reason for action. It may be a necessary background condition for my reason being a reason for me to act in the relevant way. But this is very different⁵⁰: something about me may be a part of the story of why I offer the value of autonomy as a reason for (political) action. But it's not a part of the reason itself. Again, the situation in other cases is exactly alike: Perhaps the fact that I believe that the moon's gravitational force explains the tides is a part of the background conditions needed for my offering that explanation of the tides. But this belief of mine is not a part of the explanation itself. The explanation is not about me.

Still, you may want to insist, even if nothing about me is a part of my reason for action, still I am involved in this other, backgroundish kind of way. Furthermore, even if I do not, by acting politically on the value of autonomy, give extra weight to my belief about the value of autonomy (because that belief is no part of my reason for action), still I act in a way that will, as I know, bring about a situation in which my beliefs are more in line with the state's actions than the beliefs of those rejecting the value of autonomy. In this sense too, then, it may still be argued that acting on the value of autonomy, in the face of (reasonable?) disagreement about it, violates equality. But this would be a mistake. It's true that when I act, I act on my own beliefs. There is no way around it: *anything* I do is something that *I* do. Even if I decide to defer to another, it's me doing the deferring, and indeed the deciding to defer. Even if I become a card-

⁴⁹ Raz (1998, 27) puts this point correctly but somewhat cryptically thus: "We rely on the answer to the question not because it is our view, but because it is, as we believe, true."

⁵⁰ This distinction is central to Mark Schroeder's work. See, for instance, his (2007, Chapter 2).

carrying member of the public reason club, the refusal to act politically on my comprehensive doctrine will be very much *my* refusal. My beliefs and actions are and always will be *my* beliefs and actions⁵¹. Thus, if by acting on the value of autonomy I violate equality (because it's my own beliefs I act on), then *any* action whatsoever violates equality. This is not the kind of equality worth caring about.

What the discussion in this section shows is that the equality underlying public-reason-motivation is based on a simple confusion. The liberty motivation should be, to an extent, accommodated (a point I get to below). But the equality motivation should just be rejected. The tension between authority and equality (as understood by public reason theorists) is a pseudo-problem.

5. The Role of the Political Philosopher

The Rawls of *Political Liberalism* is sometimes accused of blurring the lines between political philosophy and politics⁵²: the emphasis on stability that seems to enter the picture much earlier than it would for most others, the danger that his account is “political in the wrong way”⁵³, distancing oneself from truth and replacing it by a standard that seems to depend at least partly on what the people around us accept, the attempt to apply a principle of toleration to philosophy itself⁵⁴ – all these make the suspicion that Rawls has become a bit of a politician understandable. And Rawls and Rawlsians are of course eager to show that this isn't quite so. What I want to do in this section, though, pulls in the opposite direction. For I want to highlight

⁵¹ See Raz (1998, 27).

⁵² See, for instance, Raz (1990, 10).

⁵³ See Hampton (1989). Rawls's (1989) is motivated in large part by the need to respond to this accusation.

⁵⁴ Rawls (1993, 10).

– regardless of the details of the controversy just noted – a way in which the public reason tradition creates *too big* a divide between political philosophy and politics.

If you're a public reason theorist, how do you view those fighting in the political arena for everything they think is good and just? Perhaps some of them, of course, are themselves committed to public reason, so that they're really fighting only for what's reasonable (or some such). But much more often, the rhetoric and also sincere beliefs of political actors and activists is that of comprehensive doctrines. They have a worldview, and one that is much too rich and exciting to be acceptable according to a public reason justifiability-to requirement. With regard to these, the public reason theorist – in his capacity as a political philosopher – refuses to engage. Of course, he too may have his “private” moments, in which he engages. But when he does that, he must think of himself as “no longer engaged in political philosophy suitable for a well-ordered liberal society ...” (Quong 2011, 242). In his capacity as political philosopher, he transcends the political arena, and rather than engage the disagreements within it, merely views it as the data to be taken into account, perhaps as trouble to be contained.

There is a huge difference between thinking about a disagreement “from within”, and merely taking the fact of disagreement as something to be handled, “from the outside”. Think of any politically controversial issue about which you have a view, perhaps one you care about rather passionately, where you're rather confident that you are right: say, just in order to have a concrete example, whether the USA should intervene in the situation in Syria. And think about what it is like to engage those on the other side of this controversy. Now think about what it is like to engage in discussion (or deliberation) not over the matter at controversy itself (whether to intervene in Syria), but over how to deal with the fact that this question is controversial. This latter question too, of course, can be politically important and theoretically interesting. But

crucially, it's a *different* question⁵⁵. And just as crucially, the fact that this second question is there doesn't mean it's the *only* important question there. Surely, the first question – the first-order question about which people disagree – is also an important one to ask. But the public reason political philosopher insists that so long as it's one about which reasonable citizens may differ, it's one he – and political philosophy in general – has nothing to say about.

What is involved in thinking about the disagreement in this way, in refusing to engage the parties to the disagreement on the merits of the topic they're disagreeing about, merely thinking about how to contain the disagreement's danger and harm? Clearly, this involves thinking about the disagreement in a way that's very different from the way in which the parties to the disagreement themselves are thinking about it. For them, the crucial question is what the USA should do in response to the crisis in Syria, not what should be done *about the disagreement* about how to proceed vis-à-vis that crisis. Furthermore, it's hard to resist the feeling that there's something patronizing about the public-reason attitude to the disagreement: It's as if the public reason theorist thinks of himself as the responsible adult, viewing the children quarreling, refusing to engage the details of the quarrel, just trying to see to it that no serious harm is caused as a result of the quarrel. Respecting those engaged in the debate – indeed, respecting them as free and equal responsible adults – would here amount to

⁵⁵ There may be more questions still. One (i) is the first order question, on its own. Another question (ii) is the one in the text – about how to deal with disagreement. Possibly there's a third one – revisiting the first-order question, now with the answer to the second one in mind; roughly this is the question (iii) whether to intervene in Syria *given the disagreement*, or taking into account also the fact that people differ about (i). One doesn't have to be a public reason theorist to acknowledge a difference between (i) and (iii). Disagreement may be one normatively relevant fact here, among others. (And once we see this, we may also ask many more questions: (iv) how to respond to (iii) given that there's disagreement *about it* as well; (v) How to answer (i) given disagreement about (i) *and* (iii); (vi) how to answer (v) given disagreement about it; ...)

engaging them, not to retreating from the arena, “attempting to occupy a noncontroversial high ground”⁵⁶.

By refusing to engage in this way, the public reason political philosopher treats some of the most deeply held beliefs of those engaged in the relevant disagreement as if they were mere preferences: For with mere preferences, arguably going second-order and impartial is precisely the way to go, and certainly engaging them in argument doesn’t make much sense. Indeed, returning now to the lesson of the previous section, with mere preferences the normatively relevant fact is *that people have them*, not anything about their content directly – the reason for action of someone acting on her preference for chocolate over vanilla is the indexical one “that I prefer chocolate”. But the reason for action of someone acting in the public arena on his belief that the USA should intervene in Syria is not indexical in this way at all – it’s not about his preference for intervention, but about intervention being (as he thinks) the right way to proceed⁵⁷ (or about the normatively relevant circumstances in virtue of which it is right). The public reason theorist misses all this, and relates to such principled disagreements as if they were merely preference-based⁵⁸. And because this is most certainly not how the parties to the

⁵⁶ Not just this quote comes from Raz (1998, 47), but also the thought expressed here more widely: it echoes Raz’s insistence that treating people with respect requires engaging them, not, for instance, propagating false beliefs (1998, 43). I think that Raz overstates his case – respecting people amounts to different things in different contexts, and perhaps sometimes is best manifested by a refusal to engage them. Not, though, in the context in the text, certainly not in the patronizing public reason kind of way.

⁵⁷ For a close point see Raz (1998, 35, footnote 13).

⁵⁸ Nagel (1987, 158) does notice this, as when he says “Impartiality among persons is one thing, but impartiality among conceptions of the good is quite another.” But at the end of the day he thinks that this gap can somehow be bridged.

In another (metaethical) context, I discuss in detail the distinction between conflicts that are based on mere preferences and those that are based on moral disagreement. See chapter 2 of my *Taking Morality Seriously* (2011).

There is some similarity between the point in the text and an oft-made criticism of political liberalism, according to which it prevents (for instance) the religious from relying on their deeply held beliefs in the public domain. But the point in the text is different – it’s not about what considerations political actors are allowed to cite in the public domain, exactly, but about how the public reason theorist thinks of his own role vis-à-vis such debates.

disagreement think of their commitments and disagreement, the public reason theorist fails to respect them and their commitments. Hence the patronizing.

This means that, appearances to the contrary notwithstanding, there's a sense in which public reason theorists, with their emphasis on reasonable pluralism as a fact any acceptable political philosophy should take into account, do not take disagreement seriously *enough*. Taking the disagreement over intervention in Syria seriously means being willing to engage it, not retreating to a patronizing neutral high ground⁵⁹. Of course, even while engaging the substance of the first-order disagreement, second-order questions may be asked. And as even the non-public-reason political philosopher will agree, there are restrictions on how to engage the first-order disagreements. The fight for everything that is just and good is itself sometimes constrained by substantive considerations. Political philosophy, in other words, is hard. Surely, this should come as no surprise, to political philosophers and activists alike.

You may think that the debate has gone somewhat verbal, with public reason theorists and others talking past each other, as if simply different things are meant by the two parties when they employ such terms as "political philosophy" (this could explain also the *they-just-don't-get-it* feeling on both sides). And I'm sure that that's a part of what's going on, but it can't be the whole thing. It's not a controversy that can be eliminated by disambiguation. There is a different *vision* involved, and one that itself should be the subject of substantive discussion: For it seems to me that the good political philosopher and the political activist – at the very least, the activist who is fighting on the side of the right and the good – should be allies. Indeed, their roles are in a way continuous. True, there are differences – perhaps, for instance, there are small-scale pragmatic considerations that the political philosopher can ignore and the activist should not. Or perhaps the activist may be excused for argumentative imprecisions that cannot

⁵⁹ I make a somewhat similar point against Waldron in my (2007).

be tolerated from the political philosopher. And certainly, the political philosopher should (sometimes) engage matters on a level of abstraction that is unsuitable for activist texts. Indeed, the political philosopher may from time to time engage in finding interesting truths that have no practical bearing whatsoever - to repeat a point from section 3.3 above, that they're interesting should be reason enough for philosophical inquiry. But these differences between the political philosopher and the activist notwithstanding, the activist and the political philosopher are engaged in what is largely the same project. The public reason vision for political philosophy is different, though. On it, the political philosopher looks at activists from a distance, and with a patronizing eye. On this view, it's not just that not all political philosophy is guaranteed to be relevant for activism; rather, it's that virtually all political philosophy is guaranteed *not* to be relevant to activism. This is not what political philosophy should be like⁶⁰.

6. Accommodating (to an Extent) the Underlying Public Reason Intuitions

If I am right, then, there is no hope for a recognizably public reason view. There is no remotely plausible way of offering a necessary condition for legitimacy in terms of justifiability to the relevant constituency, in a way that remains relevant to the underlying motivations of reconciling authority with liberty and equality. But what of the intuitions that led us down this path in the first place? Think again about Estlund's Pope. Even if the thought that by subjecting the non-Catholic to the Pope's directives we're failing to treat her as an equal has been conclusively debunked (in section 5), still there's *something* problematic in so doing, isn't there? Can this fairly robust intuition be accommodated, without going all the public reason way?

⁶⁰ Nor, it seems to me, is it what it has been like in the field's long history (most of which, I confess, I do not know).

I think it can. But we should be careful about what exactly the raw, robust intuition is. The intuition is not exactly about necessary conditions for legitimacy, or about the need to justify-to, or any such thing: these are much too theoretical matters, and they come into play later, as it were, not at the level of initial intuitions. The raw intuition is the one I just stated – namely, that even if Catholicism is right, and the Pope infallible, still *there is something objectionable about subjecting a non-Catholic to the Pope's directives*. It is *this* powerful intuition that needs to be accommodated.

And it rather easily can⁶¹. For very plausibly, it's a good thing to engage people (roughly speaking) on terms that they can accept, pretty much as they actually are. Perhaps they are even pro-tanto owed to be engaged in such a way. When other things are equal, a political arrangement is better to the extent that it manages to justify itself to the relevant constituency, to the extent that its justification is accessible to its subjects pretty much given their deeply held beliefs, principles, reasons. We can give a deeper story for why this is so – I would give one in terms of autonomy, but perhaps there are other possible stories. The crucial thing is that for whatever deep reason, there's something to be said for justification-to, or something similar to it.

Note that on this suggestion nothing about justification-to is a necessary condition for legitimacy. Rather, it's yet another political desideratum, yet another item on the list of significant factors counting in favor of a political arrangement. It has no lexical priority over others, and is to be balanced against them. This suffices to fully explain Estlund's Pope example: There is indeed something objectionable about subjecting the non-Catholic to the Pope's directives, for a political order that does that is less good for doing so. Similarly for a political

⁶¹ I don't know of anyone pursuing this line, but there are some related hints in the literature. The one most clearly flirting with this line is Brower (1994, 22). And there seems to be something in this direction also in Hampton (1989, 801), Nagel (2003, 77), Estlund (2008, 163), and Kelly Sorensen's (2013) "aspirational political liberalism".

order that imposes autonomy-based directives on the members of an authoritarian illiberal community. But this doesn't mean such imposition is always and everywhere illegitimate – whether it is depends on whether it can gain in other values (or, in the latter case, in the same value, the value of autonomy) what it loses in failing a plausible justification-to requirement.

The advantages of this way of accommodating the underlying public reason intuition should at this point be clear. Because justifiability-to is no longer considered a necessary condition for legitimacy, its failure with regard to some does not entail anarchism. This means that there is no need to exclude the unreasonable, or to idealize. On the other hand, depending on the underlying, perhaps autonomy-based, story, there may be room for some gradability. Perhaps, for instance, while it's always *pro tanto* objectionable (other things being equal) to fail to justify the political authority to someone subject to it, still it's *worse* when that person is reasonable (in the natural-language sense of this term) than when she is not. Or perhaps a substantive understanding of the value of autonomy entails that choices under some mildly hypotheticalized conditions are more important for autonomy than those made under unfavorable actual conditions. If so, we get some plausible, gradable, well-motivated ways of giving weight to some of the more plausible thoughts in the vicinity here (perhaps about false consciousness, or about the drunken patient). But we do so without rendering the account vulnerable to the difficulties from section 3 above⁶².

And because no claim is being made about some spurious neutrality, and because the political philosophy that emerges from this picture is one that is avowedly based on evaluative and perhaps other normative judgments, this way of thinking of political philosophy is entirely consistent with the vision described in the previous section: The political philosopher, on this

⁶² Similarly for the self-defeat worry (again, see Wall 2002, and my 2013). And I am pretty sure that this is true also of the epistemological problems facing public reason accounts, but I have to fill in all the details there.

picture, acknowledges that all there is is the arena itself, and enters it, armed with the list of desiderata on which justifiability-to is another item, to fight, shoulder to shoulder with non-philosophers, for everything that is just and good.

I realize, of course, that many public reason theorists will not be willing to settle for this more minor role of accessibility or availability or justifiability-to. And in all honesty, the value of justifiability-to will often be outweighed by others, as the list of political desiderata is long and dense, and the stakes in politics usually high. Still, this way of accommodating the underlying public reason intuitions is the most that is directly supported by those intuitions, and more importantly, the most that can withstand criticism.

Let me not pretend that our work is done. Many more details are needed: We need an account of why it is that justifiability-to is of value; we need to understand what are the features that determine of how much value it is in specific circumstances; the relevant kind of access (or justifiability-to) has to be described, understood, and motivated; and so on. But the discussion above shows, I think, that *this* is the work that needs to be done. Nothing at all about justifiability-to is a necessary condition for legitimacy. Still, justifying political arrangements to those subject to them (in some, yet to be made precise, sense of “justifying to”) is of pro-tanto significance. When other things are equal, it should be strived for. This is all we can get, and all those tempted by public reason theory should have ever gone for.

References

Bruce Brower, (1994) "The Limits of Public Reason", *The Journal of Philosophy*, 91 (1994), 5-26.

Joshua Cohen, (2009) "Truth and Public Reason", *Philosophy and Public Affairs* 37, 2-42.

Terence Cuneo, (2013) Book Review of Paul Weithman's *Why Political Liberalism?* Faith and Philosophy 30, 357-61.

On John Rawls's Political Turn, forthcoming in *Faith and Philosophy*.

Burton Dreben, (2003) "On Rawls and Political Liberalism", in Freeman (2003), 316-346.

Ronal Dworkin, (1973) "Justice and Rights", in his *Taking rights Seriously* (Cambridge: Harvard University Press), Chap. 6.

David Enoch, (2005) "Why Idealize?", *Ethics* 115(4), 759-787.

(2007) "Taking Disagreement Seriously: Some Critical Comments on Jeremy Waldron's Law and Disagreement", *The Israel Law Review* 39, 22-35.

(2009) "On Estlund's Democratic Authority", *Iyyun* 58, 35-48.

(2009b) "More on the Conscience-Based Exemption from Military Service: A Reply to Dani Statman", 31 *Iyunei Mishpat* (the Tel Aviv Universtiy Law Review) (2009), 709-740 (in Hebrew).

(2011a) *Taking Morality Seriously* (Oxford: Oxford University Press)

(2011b) "Not Just a Truthometer: Taking Oneself Seriously (But Not Too Seriously) in Cases of Peer Disagreement", *Mind*, 119, 953-997.

(2013) "The Disorder of Public Reason: A Critical Study of Gerlad Gaus's *The Order of Public Reason*", *Ethics* 124, 1-36.

(Forthcoming) "Authority and Reason-Giving", forthcoming in *Philosophy and Phenomenological Research*.

(Manuscript) "Idealizing Still Not Off the Hook", available at law.huji.ac.il/upload/ReplyToSobel.doc

David M. Estlund, (2008) *Democratic Authority: A Philosophical Framework* (Princeton: Princeton University Press).

(2009) "Reply to Commentators", *Iyyun* 58, 73-88.

Colin Farrelly (2007), "Justice in Ideal Theory: A Refutation", *Political Studies* 55, 844-864.

Samuel R. Freeman, (ed.) (2003) *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press).

Jean Hampton, (1989) "Should Political Philosophy Be Done without Metaphysics", *Ethics* 99, 791-814.

(1993) "The Moral Commitments of Liberalism", in David Copp and Jean Hampton (eds.), *The Idea of Democracy*, 292-314.

Mark Johnston, (1989) "Dispositional Theories of Value," *Proceedings of the Aristotelian Society* 63, suppl., 139-74.

Charles Larmore, (1990) "Political Liberalism", *Political Theory* 18 (1990), 339.

(2003), "Public Reason", in Freeman (2003), 368-393.

James Lenman, (2010) "Humean Constructivism in Moral Theory", *Oxford Studies in Metaethics* 5, 175-193.

- Stephen Macedo, (1995), "Liberal Civic Education and Religious Fundamentalism: The Case of God v. John Rawls?", *Ethics* 105, 468-496.
- Thomas Nagel (1987), "Moral Conflict and Political Legitimacy", *Philosophy and Public Affairs* 26, 215-240
 (1991), *Equality and Partiality* (Oxford: Oxford University Press).
 (2003), "Rawls and liberalism", in Freeman (2003), 62-85.
- Jonathan Quong, (2011) *Liberalism without Perfectionism* (Oxford: Oxford University Press).
 (2012) "Liberalism without Perfection: Repliesto Gaus, Colborn, Chan, Bocchiola", *Philosophy and Public Issues* 2, 51-79.
 (2013) "Public Reason", in the *Stanford Encyclopedia Online*, available here: <http://plato.stanford.edu/entries/public-reason/>
- John Rawls, (1989) "The Domain of the Political and Overlapping Consensus", *NYU Law Review* 64, 233-255.
 (1993), *Political Liberalism* (New York: Columbia University Press).
- Joseph Raz, (1990) "Facing Diversity: The Case of Epistemic Abstinence", *Philosophy and Public Affairs* 19, 3-46.
 (1998) "Disagreement in Politics", *American Journal of Jurisprudence* 43 (1998), 25-52.
- Enzo Rossi, (2014) "Legitimacy, Democracy and Public Justification: Rawls' Political Liberalism vs Gaus' Justificatory Liberalism", *Res Publica* 20, 9-25.
- Mark Schroeder, (2007) *Slaves of the Passions* (Oxford: Oxford University Press).
 (2008) "Having Reasons", *Philosophical Studies* 139, 57-71.
- A. John Simmons (1999), "Justification and Legitimacy" *Ethics* 109, 739-771.
 (2010) "Ideal and Nonideal Theory", *Philosophy and Public Affairs* 38, 5-36.
- David Sobel, (2009) "Subjectivism and Idealization, *Ethics* 119, 336-352
- Kelly Sorensen (2013), "Review of Nicholas Wolterstorff *Understanding Liberal Democracy: Essays in Political Philosophy*", *Notre Dame Philosophical Reviews*, available here: <http://ndpr.nd.edu/news/42380-understanding-liberal-democracy-essays-in-political-philosophy/>
- Dani Statman (2009), "Critical Reflections on the Conscientious Exemption from Military Service", 31 *Iyunei Mishpat* (the Tel Aviv Universtiy Law Review) (2009), 669-708 (in Hebrew).
- Zofia Stemplowska (2008), "What's Ideal about Ideal Theory?", *Social Theory and Practice* 34, 319-340.

Kevin Vallier and Fred D'agostino (2013), "Public justification", in the *Stanford Encyclopedia Online*, available here: <http://plato.stanford.edu/entries/justification-public/>

Jeremy Waldron, (1987) "Theoretical Foundations of Liberalism", *The Philosophical Quarterly* 37, 127-50.

(1999) *Law and Disagreement* (Oxford: Oxford University press)

Steven Wall, (2002) "Is Public Justification Self-Defeating", *American Philosophical Quarterly* 39(4), 385-394.

(2013) "Critical Study: Public Reason and Moral Authoritarianism", *The Philosophical Quarterly* 63, 160-9.

Nicholas Wolterstorff, (2012) *Understanding Liberal Democracy: Essays in Political Philosophy*, Terence Cuneo (ed.), (Oxford: Oxford University Press).