

A Framework for Analyzing Public Reason Theories*

Paul Billingham and Anthony Taylor

A number of political philosophers are attracted to the idea that the exercise of political power ought to be acceptable to all reasonable citizens. We will refer to this requirement as the Reasonable Acceptability Principle (RAP), and to theories that endorse it as public reason views.¹ Such views can differ greatly, depending on how they specify this requirement. Nonetheless, all such views share a common structure, or so we will argue. The aim of this paper is to elucidate this common structure by identifying four questions that any public reason view must answer: the rationale question, the idealization question, the formulation question, and the content question. The answers to these four questions stand in a particular relationship to one another, which we also aim to elucidate. Indeed, we will argue that the rationale question is fundamental, in the sense that its answer should explain and justify its proponent's specification of RAP.

Part of the purpose of identifying this common structure is clarificatory. If we are to have a productive conversation about the merits of public reason views in general, and of any particular such view, then we need to understand the kinds of questions that such views must answer in order to be coherent and defensible. But our analysis also has a critical edge. Both advocates and critics of public reason views often fail to

* This is the Accepted Manuscript version of a paper published in the *European Journal of Political Theory* (published version: <https://doi.org/10.1177%2F1474885120925381>). Previous versions of this paper were presented at a Centre for the Study of Social Justice seminar and a workshop on the work of Gerald Gaus, both at the University of Oxford in 2019. We owe thanks to everyone who attended on those occasions. For comments on previous versions of the paper, thanks also to Maxime Lepoutre, the editors of this journal, and the two reviewers (one of whom revealed themselves to be Andrew Lister).

¹ We use the term 'public reason views' because it is a common and familiar way of referring to the views in which we are interested. Some might prefer other terms, such as 'political liberalism', 'justificatory liberalism', 'public justification views', and so on. But others think that there are important differences between the views picked out by those different terms. In any case, we are focused on views that take acceptability to all reasonable persons to be required for legitimacy, and use 'public reason views' to refer to these.

answer the four questions clearly, and this failure can undermine the force of their arguments. Indeed, we will argue that some of the central objections to public reason views are cast in a new, clarifying, light once we properly understand the structure of those views. This is not to suggest that no existing public reason views have the right structure, or that no objections to them successfully reflect that structure. If our analysis were wholly surprising or novel then this would cast serious doubt upon our claim to be capturing what public reason theorists are up to. Nonetheless, there is sufficient unclarity and confusion in the burgeoning public reason literature for the clarification we offer to be urgently required, and for it to generate criticisms of some familiar claims within that literature. For example, one upshot of our analysis is that it makes little sense to claim that a view ‘fails to take reasonable pluralism seriously’. This is a fairly common objection, but we will argue that it is misguided.

David Enoch (2015: 112–3) has recently expressed puzzlement with the public reason literature. He notes that many theorists seem to believe that public reason views are the only game in town, while others believe that such views have repeatedly been shown to be dead ends. Often the two sides simply seem to be talking past one another. Part of the explanation for this, we suggest, is that some versions of the view are susceptible to decisive objections because they do not have the proper structure—i.e. they fail to provide a coherent set of answers to our four questions. Critics often target these versions, and rightly point out their flaws. But those objections often do not apply to coherent versions of the view, and sometimes the objections themselves fail to properly capture the structure of such views. As a result, many public reason theorists see the objections as misguided. The mutual puzzlement that Enoch identifies is the result. We hope, therefore, that our argument in this paper can be of service to both advocates and critics of public reason views, and indeed can facilitate a more productive debate between them. It can serve advocates by encouraging them to show how their view offers a plausible and coherent set of answers to our questions. And it can serve critics by helping them to identify the precise target of their criticisms,

and to ensure that they are not simply attacking incoherent or implausible versions of the view.

The paper proceeds as follows. In the first section we explain the four questions that any public reason view must answer and the relationship between them. The following section makes a series of critical comments on claims and arguments in the literature that fail to reflect the correct structure. We highlight places where even prominent theorists in the debate make argumentative moves that our structure shows to be suspect. The third section shows that our analysis casts light on two of the most common objections to public reason views: the asymmetry objection and the self-defeat objection. This also enables us to elucidate the sense in which the rationale question is fundamental. Finally, we consider the consensus/convergence distinction, which is often treated as central in the literature, and argue that a focus on this distinction obscures much of what is important in the debate between public reason views.

The Common Structure

The first two questions within our framework both relate to the specification of RAP. A key dimension along which public reason views vary in this respect is in how they specify the constituency of reasonable citizens to whom acceptability, or justifiability, is required, which we will call a public reason view's *justificatory constituency*. On every public reason view, this justificatory constituency is an idealized constituency. The view does not hold that acceptability to all actual citizens of contemporary societies is required for law to be legitimate. And for good reason: there is little or nothing on which all actual citizens agree. But the view also does not hold that the justificatory constituency is idealized such that they simply accept the truth about what laws or policies can legitimately be enacted: that would make the appeal to acceptability to reasonable citizens redundant in any argument for the legitimacy of a particular law or policy. Apart from satisfying these two constraints, extant public

reason views offer quite significantly distinct accounts of how this justificatory constituency is to be idealized, which is sometimes expressed as a contrast between ‘moderate’ and ‘radical’ idealization (Vallier, 2014: 145–180).² Beyond the kinds of idealization that have been used by extant public reason theories, there are a plethora of possible idealizing conditions to which variants of the view could potentially appeal.

Which laws or policies a public reason view takes to be legitimate will depend fundamentally on how it idealizes its justificatory constituency. If its idealization involves taking the actual citizens of some contemporary society and asking what they would accept if their views were coherent—a very minimal kind of idealization—then there will likely still be little or nothing on which those citizens agree unanimously. This constituency could include, for example, an anarchist who denied that the exercise of political power is ever justifiable, or someone who denied that others have any moral claims at all. By contrast, idealizing conditions which ascribe a particular set of values to the justificatory constituency will naturally open up the possibility of a wider range of laws and policies being legitimate. Given the ultimate significance of which idealizing conditions are chosen for its normative upshots, one essential question we must ask when developing, presenting, or evaluating, any public reason view is what we will call the

IDEALIZATION QUESTION: How are the reasonable citizens to whom the view refers idealized? What conditions are used to specify their beliefs, desires, or commitments?

There are numerous ways in which the idealization question has been answered by proponents of public reason views, and a large set of possible ways in which it could be answered. One general distinction here is between procedural and substantive

² Vallier notes that this distinction picks out ranges within a spectrum of degrees of idealization, rather than identifying just two possible options.

conditions. Idealization could involve appealing to procedural conditions, such as what citizens would accept after going through a process of deliberating in accordance with certain norms of good reasoning (Gaus, 2011: 232–257). It could also involve imputing certain substantive values, and claims about their relative weight, to the constituency of reasonable citizens (Quong, 2011: 37–39). There are also various possible combinations of appeals to both procedural and substantive conditions.

Once a public reason view has identified its justificatory constituency, it must then explain what conditions need to be met in order for laws to be justified to this constituency. This gives us our second question:

FORMULATION QUESTION: What conditions need to be satisfied in order for an exercise of political power to be acceptable to all reasonable citizens?

There are several dimensions to this. First, there is the question of ‘scope’. Some public reason theorists hold that it is constitutional essentials and matters of basic justice that must be justified to all reasonable citizens in order for the exercise of political power to be legitimate (Rawls, 2005; Watson and Hartley, 2018: 64–72). On this view, legislation pertaining to the distribution of public funding for sports, for example, would not itself have to be acceptable to all reasonable citizens. Many defend a wider scope, holding that all laws and policies must be so justified (Quong, 2011: 256–290). Others extend this further, holding that all interpersonal moral demands must be reasonably acceptable, since the principle applies to the authority to demand compliance with moral imperatives (Gaus, 2011; Vallier, 2019). We have couched our discussion thus far in terms of ‘laws and policies’, and will continue to do so for reasons of simplicity, but everything we say also applies to views that involve a different scope.

Another dimension of the formulation question concerns whether laws are acceptable to all reasonable citizens when they are justified by appeal to a set of reasons that all

members endorse, or whether laws must themselves be endorsed by all citizens on the basis of their full set of beliefs and values. In other words, the former view here holds that it is the reasons that justify a law that must be endorsed by all reasonable citizens for that law to be acceptable to all; unshared reasons are excluded from consideration. The latter view holds that the law itself must be endorsed by all, possibly for different reasons; laws that are not unanimously endorsed must not be enacted. This dimension of the formulation question has been expressed in terms of a distinction between ‘consensus’ and ‘convergence’ public reason views (Vallier, 2011) and between the ‘reasons-for-decision’ and ‘coercion’ models (Lister, 2013: 15–23).

Importantly, both of these kinds of views face further questions of specification. The ‘consensus’ or ‘reasons-for-decision’ view must explain what it means to say that laws are justified by the values shared by reasonable citizens. Does this require that all citizens agree that those reasons support the law, or is it sufficient that there is some plausible argument for the law that draws upon those reasons?³ If the latter, then what is the standard for ‘plausibility’ here?

Turning to ‘convergence’ or ‘coercion’ views, we need to know what it means for all reasonable citizens to ‘endorse’ a law. Does this mean that they must consider the law optimal—i.e. that it is their most preferred law within the relevant policy area? Or is it enough for them to consider the law to be an improvement against a baseline of no law in this area? Or an improvement against the status quo, or some other baseline? Again, which laws or policies a public reason view takes to be legitimate will be shaped by its answer to the formulation question, since some of the answers we have suggested are much more restrictive than others. It is thus crucial for public reasons theorists to make clear their answer to this question.

³ Another way to express this distinction is to use Vallier’s (2014: 108–111) contrast between ‘shareable’ and ‘accessible’ reasons.

Any particular answer to the idealization and formulation questions stands in need of defense. It would be unacceptable for a public reason view to pick its idealizing conditions arbitrarily, so a rationale will have to be given for the decision to idealize one way rather than another. Similarly, an explanation is needed for any particular formulation of RAP. This leads us to the third question that we must ask when developing or evaluating any public reason view, namely the

RATIONALE QUESTION: Why should we endorse this version of the Reasonable Acceptability Principle? Why does the legitimacy of laws and policies depend on what this constituency of reasonable citizens would or would not accept?

In order to be successful, an answer to the rationale question needs to explain and justify its proponents' answers to the idealization and formulation questions. It needs to tell us why we ought to care about justifiability to the constituency specified by that answer; and, more specifically, why we ought to tie the legitimacy of laws and policies to what could be justified to this constituency. It also needs to tell us what needs to be acceptable to this constituency, and what the requirement of acceptability amounts to. In this sense it is the most important question the view needs to answer.

So far, we have argued that all public reason views require an answer to the idealization, formulation, and rationale questions. Before turning to the final element of our framework, it is worth considering a potential challenge to the claim that such views must offer an answer to the rationale question. It may seem obvious that any defense of a public reason view must explain why we ought to endorse its version of RAP. However, the rationale question is not always given due consideration in the literature, and there is a ready explanation of this fact. In his early work in this area, Rawls stated that his aim was to 'try, so far as [possible], to avoid disputed philosophical, as well as disputed moral and religious, questions' (Rawls, 1985: 230), and that his view was formulated by 'apply[ing] the principle of toleration to philosophy itself' (Rawls, 1985: 223). In this description of his philosophical project,

Rawls likely encouraged the avoidance of questions about the foundations of public reason views. Proponents of public reason have avoided discussion of what the rationale for RAP is, so the thought goes, because the possible rationales to which they might appeal will be disputed philosophical positions of the kind that a public reason view should seek to avoid.⁴

The rationale question must be confronted head on, however. Public reason theorists need to explain why we ought to accept a public reason view, and indeed their particular public reason view. To not answer this question would be to hold, absurdly, that a controversial philosophical view does not stand in need of defense. This is especially clear once we see that there are a variety of potential ways in which RAP could be specified; we need a reason to choose one specification over another.

A number of distinct answers to the rationale question have in fact been proposed. Rawls and some of his followers suggest that adherence to RAP is necessary in order for a society to be stable for the right reasons (Weithman, 2010). Charles Larmore (1999), Thomas Nagel (1987), and Martha Nussbaum (2011) argue that the principle is best understood as grounded in a norm of equal respect for persons. Other proffered rationales include the realization of an ideal of civic friendship (Lister, 2013a), political autonomy (Weithman, 2017), justice (Quong, 2013), reciprocity (Neufeld, 2010), and the rational sustainability of our reactive attitudes (Gaus, 2011). Each of these values has been proposed as a rationale for RAP.

As well as its answers to the idealization, formulation, and rationale questions, the defensibility of any public reason view will also depend on its answer to the

⁴ One clear manifestation of this thought is Charles Larmore's (1990: 342) search for a rationale for a principle of liberal neutrality that is itself neutral, or as neutral as possible. Another is Leif Wenar's (1995) view that Rawls ought to jettison aspects of his account that plausibly constitute his answer to the rationale question—such as his political constructivism—on the basis that they are likely to be controversial among reasonable citizens.

CONTENT QUESTION: What, if anything, is acceptable to all reasonable citizens? What laws and policies does this view deem to be legitimate?

Public reason views differ dramatically with regard to the kinds of laws and policies that they conclude to be legitimate. Their answer to the content question is a function of their answers to the idealization and formulation questions. What can be justified to all reasonable citizens will depend on how those citizens are specified—and in particular what reasons, values, or ideals those citizens endorse—and what conditions need to be satisfied for laws to be acceptable to those citizens. As we have seen, one's answer to the idealization and formulation questions must itself be justified by one's answer to the rationale question. So in that sense, the practical differences with regard to what laws and policies different public reasons view deem to be legitimate are ultimately rooted in their different rationales.

Public reason theorists clearly need not spell out all of the implications of their view or seek to determine precisely what set of laws it deems legitimate, especially since this might well depend on various empirical facts about particular societies. Nonetheless, in order to evaluate any public reason view, we will need to consider whether we think that its implications are at least minimally acceptable after due reflection. Numerous objections to public reason views attack them for what are deemed to be unacceptable conclusions about which laws and policies are legitimate. One such objection holds that a particular public reason view entails anarchism, since there are no laws or policies that are unanimously acceptable to its constituency. Others have sought to establish that there are particular important issues for which the standard set by the public reason view can provide no answer, even if it does not entail anarchism.⁵ Alternatively, objectors have sought to show that there is a particular issue on which a public reason view necessarily delivers an unpalatable

⁵ This is the so-called 'incompleteness objection'. For a recent variant, see Kramer (2017: 92–155). For earlier versions, see Greenawalt (1988), Reidy (2000), and the discussion in Schwartzman (2004).

answer. For example, Jeremy Williams argues that public reason has unacceptably permissive implications for abortion, permitting termination ‘with little or no qualification, right until birth’ (2015: 25).

While it is fairly uncontroversial that it would be problematic for a public reason view to conclude that there are no legitimate laws, the question of when we ought to revise or reject a view because it delivers an intuitively unpalatable result on a particular issue is less straightforward. It is part and parcel of the method of reflective equilibrium that we should be open to changing the principles to which we are committed if they cannot be made to cohere with our considered convictions about particular cases. Contrariwise, it is also part and parcel of that method that we should be open to modifying our judgments about particular cases in order to make them cohere with principles that we find independently compelling and that explain our convictions about other cases. We cannot state any general rule in advance that specifies precisely when we ought to make such modifications in one direction rather than the other. However, this does not tell against the claim that the proper evaluation of a public reason view must include a consideration of its answer to the content question.⁶

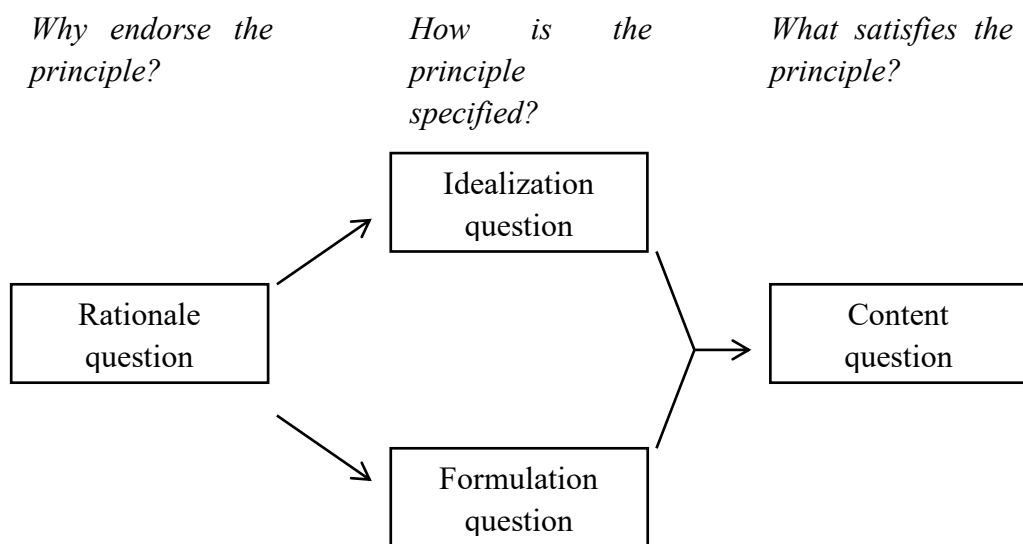
In sum, then, the common structure of public reason views is given by their answers to the idealization question, the formulation question, the rationale question, and the content question.⁷ Of these four questions, the rationale question is the most important. An answer to that question needs to explain why we ought to tie the legitimacy of law to what is acceptable to all reasonable citizens. Moreover, it needs

⁶ In her recent *Liberalism, Neutrality, and the Gendered Division of Labour* (2019), Gina Schouten expresses a complimentary view about the significance of the content question. She notes that her willingness to endorse political or public reason liberalism is to some extent contingent on the view’s compatibility with certain gender egalitarian policies (14), but also that the view need does not need to endorse every commitment of gender egalitarianism in order to be attractive (160).

⁷ One issue we have not mentioned is who (if anyone) is under duties to present public reasons within their political advocacy. For some theorists, this is a separate question from the specification of RAP, while for others it may fall under the content question (and thus similar comments would apply to those we made about that question).

to explain why we ought to tie the legitimacy of law to what is acceptable to all reasonable citizens idealized in the manner given by the answer to the idealization question. And it needs to explain what it means for laws to be acceptable to those citizens. Of course, this is not to downplay the significance of the content question. While the answer to this question is a function of one's answer to the previous questions, an implausible or unacceptable answer to this final question gives strong reason to revise the overall view, as we have discussed.⁸

This structure is shown in the diagram below.



We will encounter several views that display the formal structure that we have elucidated in the following sections, where we apply our framework to various important disputes within the literature. Before doing so, a note of clarification about the notion of reasonable disagreement is necessary. Within public reason views, this notion is tied to the principle that the exercise of political power is legitimate only if it

⁸ Some may worry that our framework objectionably implies that there is a 'truth' about how RAP should be specified. This might be thought objectionable because public reason views (or at least Rawlsian versions) are meant to make no appeal to the idea of truth. However, as Quong (2011: 224) argues, even the Rawlsian view cannot do without truth in the 'mundane' sense, where to say that something is true is just to say that it is 'correct, right, valid, or sound'. Public reason theorists should only wish to avoid appeals to metaphysical claims about what ultimately makes propositions true. Our framework is compatible with this kind of truth-abstinence.

is acceptable to all reasonable citizens. However, the term ‘reasonable disagreement’ has become widespread within philosophical parlance since public reason views were introduced into the literature, and not all uses of the term imply that the speaker accepts RAP (see Laborde, 2017; Valentini, 2013). Some use the term simply to refer to a domain of issues that are controversial among well-meaning people. On this usage, the fact that an issue is the object of reasonable disagreement does not have any direct implications for the legitimacy of laws. In setting out the common structure of public reason views, we mean to identify the questions that proponents of RAP must answer, rather than all those who make use of the term ‘reasonable disagreement’.

Applying the Framework I: Clarifying the Debate

In this section we will apply our framework to several common ideas within the debate concerning public reason views, in order to show that some claims that are used to criticize and defend such views fail due to misunderstanding or misrepresenting the structure of those views.

It is common for people to criticize particular public reason views on the grounds that they fail to take reasonable disagreement seriously, or that they underestimate the extent of reasonable pluralism (Freyenhagen, 2011; Vallier, 2014: 121, 158–160). Our framework reveals an important problem with such claims. The term ‘reasonable disagreement’ refers to disagreement among members of the justificatory constituency. There is thus no answer as to what ‘reasonable disagreement’ *is* without an answer to the idealization and rationale questions. ‘Reasonable disagreement’ is not something that is out there in the world, and to which public reason views must respond. It is a construct of public reason views themselves.⁹

⁹ Quong (2011: 138–145) distinguishes two versions of the public reason view—the internal and external conceptions—partly based on whether they take reasonable pluralism to be a ‘fact about the world’ or a ‘fact about liberal theory itself’. He presses a number of objections against the external conception, one of which is that if it appeals to idealizing conditions then it collapses into the internal conception. In one sense, our point here is a generalization and extension of this claim. However, Quong tends to

If a particular public reason view has an answer to these questions that leads to a particular specification of the justificatory constituency then one cannot coherently object that this specification is one that ‘fails to take reasonable disagreement seriously’.¹⁰ Reasonable disagreement, according to this public reason view, simply *is* the disagreement exhibited by members of its justificatory constituency (Billingham, 2017: 552–553). And the advocate of the view can rightly insist that they do take reasonable disagreement seriously; after all, they hold that laws are legitimate only if they are acceptable to this constituency. The critic here must be presupposing their own answer to the idealization question, according to which members of the justificatory constituency disagree in a more wide-ranging way. From the perspective of *that* idealization the view they are criticizing seems to underestimate the extent of reasonable disagreement, or not to take it seriously. But the force of the objection must come from showing that this alternative idealization is normatively superior—that the justificatory constituency that we ought to care about is the one specified in this way rather than that (Billingham, 2017: 557). And this will be the case only if it is an implication of the most plausible answer to the rationale question. In the end, then, the real question here is which rationale (and thus which idealization) we ought to accept. One theorist might think that another has an implausible rationale, which leads to an overly-homogenous justificatory constituency. But the objection is then not that that view underestimates the extent of reasonable disagreement but that it has a different, and incorrect, conception of reasonable disagreement.

This also means that reasonable disagreement cannot itself be the answer to the rationale question. The rationale for public reason views cannot simply be that we see

speak as if there is one version of the internal conception (his own view), whereas on our account all public reason views are ‘internal’ in the sense that what counts as ‘reasonable disagreement’ is defined internally to the theory.

¹⁰ One could, however, legitimately criticise a theorist for underestimating the extent of disagreement among the justificatory constituency identified by their own theory.

something called reasonable disagreement in the real world, since reasonable disagreement is a construct of those views themselves. Further, the rationale also cannot simply be that intelligent and well-meaning citizens tend to disagree about many matters: religion, the nature of the good life, justice, and so on. It is clearly true that intelligent and morally motivated people disagree in these ways, but it does not follow from this that we ought only to exercise political power in ways that intelligent and well-meaning citizens could be expected to accept. Moving from the descriptive claim that persons with certain features tend to disagree to the normative claim that this disagreement constrains the legitimate exercise of political power will always require further argument. Whatever way the idealization question is answered, the rationale for RAP can never be the phenomenon of reasonable disagreement (among persons with certain features) itself. An answer to the rationale question needs to explain why it *matters* that persons with certain features tend to disagree. And it is that disagreement, among those persons, that then constitutes ‘reasonable disagreement’, according to the public reason view that endorses that rationale.¹¹

Another idea associated with public reason views that our framework shows cannot function as an answer to the rationale question is stability. Some interpreters of Rawls have taken his concern with stability to be about what we might call ‘empirical stability’: a lack of violence, the survival of extant political institutions, and so on (Klosko, 1993). Critics rightly point out that this concern is an implausible rationale for a public reason view, since empirical stability of this kind is unlikely to be dependent on whether laws are acceptable to members of an idealized constituency (Klosko, 1994). As many others have highlighted, however, Rawls is in fact concerned

¹¹ This does not mean that real world disagreement is entirely irrelevant for public reason theories. Observations about disagreement in the world as it is will rightly influence our judgments about what disagreements would persist under certain idealizing conditions. For example, Rawls’s reasonable citizens are members of a well-ordered society, and judgments about what citizens of well-ordered society would and would not accept will depend on various facts about the real world. See Weithman’s (2010: 158–163) discussion of ‘balance conditionals’.

with ‘stability for the right reasons’. This stability is defined as being realized when citizens of the well-ordered society freely and reflectively endorse the conception of justice that governs and regulates the basic structure (Weithman, 2010). But stability for the right reasons cannot be an answer to the rationale question either, because it amounts to a restatement of Rawls’s public reason view itself. (For a similar thought, see Mulhall and Swift, 1992: 186.) Rawls’s answer to the idealization question is that the citizens to whom justification is required are the citizens of a society well-ordered according to political liberal principles. Stability for the right reasons therefore cannot tell us *why* our principles must be acceptable to the citizens of a well-ordered society. The property of being stable for the right reasons *just is* the property of being acceptable to that idealized constituency. The claim that principles must be stable for the right reasons and the claim that they must be justified to citizens of the well-ordered society are identical, so the former cannot do the work of explaining why we should endorse the latter. An appeal to this kind of stability cannot answer the rationale question. This is not to say that stability for the right reasons cannot play any role within a public reason view. We might have some further reason to care about such stability—such that it must be achieved if citizens are to realize political autonomy (Weithman, 2010: 288–292), or that it is a desideratum that any normative theory must meet (Krasnoff, 1998). But the answer to the rationale question would then come from this further reason or value, rather than from stability itself.

As we argued above, an answer to the rationale question both explains one’s answer to the idealization question and *justifies* that idealization, by justifying the claim that acceptability to persons idealized in this way matters for legitimacy. This enables public reason theorists to respond to one of Enoch’s central objections. Enoch argues that public reason views cannot properly motivate their idealization, and that idealization is therefore an ad hoc maneuver to avoid unwelcome implications such as anarchism. This objection relies on a particular, rather vague, answer to the rationale question, however. Enoch (2015: 115) sketches what he takes to be the central

underlying public reason intuition as ‘the attempt to reconcile authority with liberty.’ This explains why exercises of political power must be justifiable *to* each person subject to them. ‘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’ (Enoch, 2015: 115). Enoch recognizes that his characterization of this thought is ‘somewhat loose,’ but claims that it is sufficient to make his objection to idealization successful.

For an objection to idealization to succeed, it must show that the idealization involved in a particular public reason view is inconsistent with that view’s own underlying rationale. Enoch stacks the deck against public reason views by claiming that they all share the underlying rationale that he identifies—a rationale that explicitly seeks acceptability to all actual citizens ‘pretty much as they are’. Such a rationale clearly places severe limits on the kinds of idealization that could coherently be invoked. Based on *this* rationale there is no good justification for a construal of ‘reasonableness’ that ‘excludes’ many real-life citizens. A public reason view that answered the rationale question in the way that Enoch presupposes would have to invoke moderate forms of idealization, at most. To that extent, Enoch is correct. But this cannot function as an objection to all public reason views, as he claims.¹² All it shows is that views that include more ‘radical’ forms of idealization must provide a rationale that justifies such

¹² Enoch’s misstep here is perhaps explained by the fact there is a trivial sense in which it is correct to claim that all public reason views aim to reconcile the freedom and equality of citizens with political authority. Most proponents of public reason couch their arguments as providing a reconciliation of freedom and authority. But they disagree about which conception of freedom and equality ought to be reconciled with political authority, and their disagreement about this matter is—so far as we can tell—simply a restatement of their disagreement about how the idealization question should be answered. An example may make this point clearer: according to Gaus’s (2011: 19) ‘restricted’ view of freedom and equality, ‘to respect others as free and equal moral persons is to refrain from claiming moral authority over them to demand that they do what they do not themselves have reason to endorse’. On this view, the way we reconcile the freedom and equality of persons with political authority is therefore by applying Gaus’s principle of public justification—his version of RAP. In this way, the restricted view of freedom and equality is little more than another way of stating RAP itself, and therefore could not be Gaus’s answer to the rationale question. Though we do not have the space to demonstrate it here, our view is that this is also what is going on when other proponents of public reason appeal to the idea of reconciling freedom and authority.

a specification of the justificatory constituency. Certainly, views that endorse Enoch's putative rationale and yet invoke a highly idealized justificatory constituency fall prey to his objection. And there might well be extant public reason views that are susceptible to an objection of this form. But not all are. It is not difficult to see that a different rationale could justify the kinds of idealization that Enoch criticizes. And many extant public reason views do offer an account of idealization that coheres with, and is well-motivated by, their rationale.

The need for a public reason view's account of idealization to cohere with, and indeed be motivated by, its rationale also gives rise to another critical point, this time concerning one of the foremost advocates of such a view, Jonathan Quong. Quong (2011: 37–39) offers a fairly clear account of his answer to the idealization question: reasonable citizens are those who are committed to the moral ideal that persons are free and equal, and to the ideal of society as a fair system of cooperation, and therefore accept and prioritize the values of freedom, equality, and fairness that are embodied in those ideals. Quong (2011: 161–255) uses this account as a key tool in responding to various objections to public reason views. His answer to the rationale question, however—which, as we have seen, ought to justify his account of idealization—is surprisingly unclear. In his *Liberalism Without Perfection* he gestures toward a stability-based answer to the rationale question (Quong, 2011: 158). But, as we have already argued, this cannot succeed. He also suggests that the rationale is to show that liberalism is practically possible, which is achieved by showing that liberal principles can be endorsed by this idealized constituency (Quong, 2011: 159–160). It is unclear why the idea of practical possibility would give us a reason to choose public reason views over alternative accounts of legitimacy, however, let alone why it would uniquely pick out Quong's particular account of idealization (see Jensen, 2009). More recently, Quong (2013: 273–275) has suggested a justice-based rationale, according to which laws being justified to reasonable citizens is a requirement of justice. Quong has only provided a brief sketch of this argument, so it is not clear how this rationale

works. The important point for our purposes is that it seems surprising that Quong can have such a clear answer to the idealization question—and, indeed, the formulation and content questions—while adopting such an underdeveloped answer to the rationale question. Further, it is unclear how Quong can experiment with different rationales while keeping his account of idealization fixed. This way of proceeding seems to assume that one’s account of idealization is somehow independent of one’s rationale. Yet, as we have argued, the former in fact must depend on the latter.¹³

Applying the Framework II: Asymmetry and Self-Defeat

As the previous section illustrated, a virtue of our framework is that it casts clarifying light on several central debates within the public reason literature. In this section we show that this is also the case with regard to two of the most common and important objections to public reason views.

One of the main objections that has been pressed against public reason views notes that it is far from obvious that reasonable citizens agree about justice but disagree about the good. Proponents of this ‘asymmetry objection’ argue that there is just as much disagreement about questions of the right as questions of the good. The claim that the state must be neutral between conceptions of the good is thus ‘unacceptable because it also issues in the unacceptable conclusion that the state ought to abstain from enacting fair and equitable principles of justice’ (Caney, 1998: 19; see also Waldron, 1999: 149–163).

¹³ Similar comments apply to R.J. Leland’s (2019) recent defence of public reason by appeal to civic friendship. A notable feature of his discussion is that he seems to assume that one’s account of idealization is independent from one’s choice of rationale (see 82), which, as we have argued, is not the case.

Proponents of the asymmetry objection rarely clarify who they have in mind by 'reasonable citizens'. The objection would certainly succeed if 'reasonable citizens' were simply well-meaning persons. It is plausible that there is very little that such persons agree upon, or that is acceptable to all such persons. As we have seen, however, public reason views rarely involve such an unidealized constituency. Instead, they hold that laws must be acceptable to a justificatory constituency that is idealized in various ways. The members of this constituency might well agree on various ideals or principles of justice, while holding diverse conceptions of the good. Thus, Quong (2011: 192–220) argues that reasonable citizens' disagreements about justice are always 'justificatory', meaning that the parties to the disagreement share basic values and ideals but disagree on their interpretation and implications. Their disagreements about the good, meanwhile, are likely to be 'foundational', meaning that the parties share no premises that can serve as a mutually accepted standard of justification. This asymmetry thus explains why laws justified by appeal to relevant principles of justice are legitimate while those justified by appeal to particular conceptions of the good are not (though, for critical discussion, see Fowler and Stemplowska, 2015; Laborde, 2017: 92–110).

Of course, this reply to the asymmetry objection relies upon providing an answer to the rationale question that explains why members of the justificatory constituency display these particular agreements and disagreements. Whether there is in fact an unwarranted asymmetry with regard to the treatment of justice and the good within a public reason view depends on whether its rationale generates a justificatory constituency that agrees on certain matters of justice but disagrees on the good, such that the answer to the content question is that laws promoting justice are acceptable to all members of the justificatory constituency, while those promoting the good are not. As we have seen, Quong does not provide a clear rationale for his idealization, which leaves his response to the asymmetry objection ultimately undefended. But if an appropriate rationale were offered then this would defuse the objection. Successful

responses to the asymmetry objection thus require successfully answering the four questions involved in our framework, while public reason views that fail to do this might well generate an objectionable asymmetry and thus be vulnerable to the objection. The correct use of our framework is thus the key to defusing this prominent objection, if it is possible to defuse it.

This point also holds for another common objection, which is that public reason views are self-defeating (Wall, 2002; Raz, 1998). This objection begins by noting that the justification of laws within public reason views will involve appeal to the public reason view itself. The fact that a law is acceptable to all members of the justificatory constituency is what makes a law legitimate. But the public reason view might itself be rejected by some members of that constituency. If so, then the law will not be legitimate after all, since its justification will not be acceptable to all members of the justificatory constituency. The view would therefore be self-defeating. Further, advocates of this objection insist, it is highly plausible that some reasonable citizens will reject RAP, since it is very controversial. In other words, it is likely that some agents who a public reason view identifies as part of its justificatory constituency reject the view itself, rendering it self-defeating.

One simple way to defuse this objection is to deny that any reasonable citizens reject one's public reason view. In other words, one can insist that all members of the justificatory constituency accept that acceptance to this constituency is legitimacy-conferring. Several theorists have made this claim, including Quong (2011: 38–39), David Estlund (2009: 55), and Andrew Lister (2013: 127).

To many, this move might seem objectionably ad hoc. And it could be. But it will not be ad hoc if it is properly motivated by one's rationale. Some rationales could lead to an answer to the idealization question that includes endorsement of the view itself as

one of the features of members of the justificatory constituency. Such a view would not be self-defeating.

Lister's view is a good example here. Lister's (2013: 105–133) answer to the rationale question is an appeal to the ideal of civic friendship. He argues that in order for citizens to enjoy this valuable form of relationship they must justify laws to one another by appeal to reasons that they all accept. 'Public reason makes possible civic friendship despite deep disagreement' (Lister 2013: 105; for critical discussion see Billingham, 2016). Such friendship depends upon reciprocity, however; individuals can only enjoy civic friendship with others who are also willing to appeal to public reasons within their political advocacy. Further, there is a cost to complying with the requirement to offer public reasons: it might well mean that individuals cannot appeal to what they take to be the truth about morality or justice. Given both of these points, individuals are only required to offer reasons acceptable to others who are also willing to offer mutually acceptable reasons. After all, the costs of compliance with public reason are only justified if such compliance produces the benefits of civic friendship, and this is only the case with respect to others who are also willing to bear those costs by reciprocating in the offering of mutually acceptable reasons. Thus, the content of public reason is determined with reference to the reasons that are acceptable to citizens who endorse the ideal of public reason itself. In other words, only those who endorse that ideal are within the justificatory constituency. Given the civic friendship rationale, it is natural for idealization to include acceptance of the requirement to offer mutually acceptable reasons for laws.¹⁴ And this means that the view is not self-defeating.

Something similar can be said regarding the other main line that public reason theorists have taken in response to the self-defeat objection, which is to deny that their

¹⁴ Lister (2018: 78–83) makes this argument himself, in response to Enoch. See also Lister (2013: 121–128).

public reason view applies to itself. Gerald Gaus (2011: 225–228) makes this move via an analogy with the falsification principle in science (see also Lister, 2018: 71–74). The principle that an argument must contain a falsifiable empirical premise in order to be admitted into scientific discourse need not be applied reflexively; it need not itself contain a falsifiable empirical premise in order to be admitted into scientific discourse. Gaus argues that we should think of his public justification requirement in the same way. Whether or not this move is legitimate depends on the answer to the rationale question (as Enoch, 2013, also points out). The claim that a public reason view does not apply to itself will be objectionably ad hoc if it is not motivated by the view’s answer to the rationale question. The reply succeeds only if the view’s rationale justifies RAP as a meta-principle, analogous to the falsification principle, which is not itself used as a justification for laws.

An upshot of our analysis here is that the success of a public reason view’s response to both the asymmetry objection and the self-defeat objection ultimately depends on its answer to the rationale question. Whether a view’s answer to the idealization question can plausibly specify the justificatory constituency such that all its members accept certain values or ideals—such as principles of justice, and the ideal of public reason itself—while disagreeing on others matters—such as conceptions of the good—depends on whether this form of idealization is well-motivated by the view’s rationale. While the self-defeat and asymmetry objections will be successful against some—perhaps many—public reason views, it is far too quick to suppose that public reason views *must* be self-defeating because it is implausible to think that all reasonable citizens endorse public reason, or that such views must fail because it is implausible to think that there is an asymmetry with regard to reasonable citizens’ (dis)agreement on matters of justice and matters of the good. Whether these claims can be sustained depends on the idealization involved in specifying the constituency of ‘reasonable citizens’. And whether any particular idealization is justifiable depends on whether we have been provided with a rationale that explains and justifies it. The

fundamental question regarding the plausibility of any public reason view is thus whether its rationale is persuasive, in terms of showing why we ought to tie a moral property such as legitimacy to the fact of unanimous acceptability among a particular justificatory constituency. There are two parts to this. First, does the rationale motivate this particular kind of idealization? Second, is the rationale persuasive? Have we been provided with good reason to think that respect for persons or civic friendship or political autonomy or the rationality of our reactive attitudes require that laws are acceptable to all members of this constituency?¹⁵

The fact that these are the fundamental questions is not surprising, once we recognize the structure of public reason views. Much discussion in the literature misses or obscures these questions, however, due to a lack of clarity regarding this structure. As should be clear, our claim here is not that objections such as asymmetry and self-defeat can never be successful; in many cases they will be. Instead, our claim is that in order to be successful they must have the right target: they must show that the rationale invoked by a particular public reason view does not justify the idealization and formulation that that view requires if it is to justify asymmetry and avoid self-defeat. In other words, they must show that there is a lack of alignment between the view's answers to the questions in our framework. Further, it is an implication of our argument that the success (or otherwise) of these objections crucially depends on the particularity of the relevant public reason view. They cannot be successful or unsuccessful against public reason views 'in general'. Indeed, even if it turned out that every possible public reason view fell foul of these objections, this would be due to the objections succeeding against each view individually, rather than due to any general or structural problems with such views. Proponents of these objections thus must target them at particular public reason views. In practice, they have often failed to do so.

¹⁵ Where 'acceptable to' is understood in particular way, as specified by the answer to the formulation question.

Consensus vs. Convergence

As the discussion so far has shown, prominent defenders and critics of public reason views have been guilty of misapplying or ignoring the common structure that we have defended. Perhaps, however, they are simply applying different distinctions to the debate, which better capture what is at stake. In this section, we consider a prominent alternative way of dividing up public reason views and argue that our common structure better captures the differences between such views, thus providing a more illuminating lens through which to look at the debate.

It is fairly common to focus on the distinction between consensus and convergence models as the main dividing line between public reason views. The views developed by Gaus and Vallier are paradigmatic examples of the convergence model, whereas Rawlsian variants are exemplars of the consensus model. The debate between these two accounts of public reason is therefore often framed as a debate between the consensus and convergence models.

According to Lister (2018), the key import of the consensus/convergence distinction is that the main objections to public reason views apply differently across it.¹⁶ The consensus model holds that that it is the reasons behind our political decisions that must be endorsed by all reasonable citizens. Reasons that are the object of disagreement among reasonable citizens are excluded from deliberation. On the convergence model, by contrast, it is laws themselves that must be endorsed by all reasonable citizens, possibly for different reasons. There is a presumption against state action that can only be defeated by agreement about the law among reasonable

¹⁶ Our construal of the distinction here follows Lister, but we use the more well-known ‘consensus’/‘convergence’ terminology, whereas he speaks of the ‘reasons-for-decision’ and ‘coercion’ models, as we noted above. We treat these distinctions as equivalent. The only difference between them that Lister (2018: 70, fn. 11) notes is that consensus is seen as involving ‘double unanimity’. But no extant consensus view exhibits this double unanimity, as far as we know. Certainly Rawls’s view doesn’t, for example.

citizens. Lister argues that the convergence variant is not vulnerable to the self-defeat objection because it does not apply to itself—it applies to coercive laws, but is not itself a coercive law. It is, however, potentially vulnerable to the objection that it entails anarchism: there might be no laws that are endorsed by all reasonable citizens. The reverse is true of the consensus model. This model does apply to itself—as it will inevitably be appealed to in the justification of particular laws and policies—so the threat of self-defeat looms. But the anarchism objection is not an issue. Even if the justificatory constituency is only weakly idealized, this just means there will be fewer public reasons, and ‘fewer public reasons means a simpler cognitive task, in deciding where the balance of public reasons lies, not a presumption against enforcing common rules’ (Lister, 2018: 70).

We agree that the choice between these two models is significant—it falls under our formulation question, as we noted above—and we agree with much of Lister’s argument. However, to focus solely on this choice would be to miss much of what is at stake. Take the claim that the consensus model is not vulnerable to the anarchism objection. This is only true if we take as fixed certain answers to the idealization question. Even on this model, as we decrease the level of idealization we approach a point at which *all* reasons are excluded from our deliberations, as none are endorsed by all members of the justificatory constituency. At this point, even the consensus model will result in anarchism. Lister might reply that the consensus model does not in fact have anarchistic implications in this scenario. When all reasons are excluded from our deliberation the model might instead hold that—because there are no reasons that speak in favor of repealing our current laws—no legal change would be acceptable. But this implication is no less damning, for the public reason view would then imply an extreme and implausible kind of status quo bias. Therefore, the consensus model’s capacity to avoid this objection is dependent on a particular answer to the idealization question. As for the convergence model, while it is true that its formulation enables it to avoid the self-defeat objection, whether it is vulnerable to

the anarchism objection also depends on the level of idealization that is employed: a convergence model with a sufficiently idealized constituency could avoid this objection. We therefore hold that discussions about which version of the public reason view to accept (if any) would be more productive if they focused on the rationale, idealization, formulation, and content questions, while acknowledging the priority of the rationale question, rather than exclusively on the consensus/convergence distinction.

Another example helps to make the same general point. The differences between Rawls's and Gaus's answers to the content question are driven as much (if not more) by other factors than the fact that the former is a consensus view and the latter a convergence view. Take the 'classical liberal tilt' of Gaus's theory: his claim that egalitarian redistributive principles often will not satisfy RAP (2011: 263). A key feature of his view that generates this conclusion is its moderate level of idealization, as it is this that ensures that there are classical liberals within the justificatory constituency. Further, Gaus's answer to the formulation question is not simply 'convergence', but a specific understanding of the baseline against which proposed laws are measured ('no law'), and the use of this baseline is also crucial to his argument that property rights satisfy RAP but egalitarian policies do not. As for Rawls, his model could also deliver Gaus's classical liberal tilt, if it used a lower level of idealization. Again, then, focusing solely on the consensus/convergence distinction would lead us to miss much of what is at stake here.

Conclusion

In this paper we have argued that in order to be defensible any public reason view will need to provide answers to the idealization question, the formulation question, the rationale question, and the content question. Answers to the idealization and formulation questions tell us what RAP amounts to—what it would take to satisfy it. The answer to the rationale question should explain and justify the answers to these

two questions, telling us why we ought to accept this version of the principle. And the answer to the content question will then be a function of the answers to the other three questions. These four questions and the relationship between them constitute a common structure onto which existing public reason views can be mapped and assessed. Explicit acknowledgement of this structure in the literature is long overdue. As we have shown, a failure to keep it front and center when thinking about these views has had a detrimental impact on the philosophical debate.