The Dark Knowledge Problem: Why Public Justifications are Not Arguments

Sean Donahue
School of Philosophy, Australian National University, Canberra, Australia
Sean.Donahue@anu.edu.au or scd.donahue@gmail.com

Abstract

According to the Public Justification Principle, legitimate laws must be justifiable to all reasonable citizens. Proponents of this principle assume that its satisfaction requires speakers to offer justifications that are representable as arguments that feature premises which reasonable listeners would accept. I develop the concept of dark knowledge to show that this assumption is false. Laws are often justified on the basis of premises that many reasonable listeners know, even though they would reject these premises on the basis of the further considerations that speakers implicitly rely on for their support. Accommodating the fact of dark knowledge requires us to consider the civic virtue of speakers to be more important for public justification than the acceptability of their arguments to reasonable citizens. I sketch an alternative conception of public justification that incorporates these results and argue that it provides a rationale for ignoring the otherwise sound contributions of some participants in political deliberation.

Keywords

civic virtue – political epistemology – public justification – public reason – testimony

A prevalent idea within contemporary liberal philosophy is that we ought to respect other citizens as free and equal in their capacity to reason in our use of democratic power.[1] This requires enforcing laws while leaving reasonable citizens equally free to live by their own best lights, even if we consider their resulting choices to be eccentric, bizarre, or even wrong (provided they do not cause excessive harm). Although some of us may be more intelligent or better informed, this does not permit us to force other citizens to live according to laws regarding which they cannot see reason to accept.

For anyone who considers this idea plausible, the philosophical tradition of Public Reason Liberalism offers a problem and solution. The problem arises from the so-called Fact of Reasonable Pluralism. Reasonable people who are at least moderately free naturally develop a plurality of incompatible views about the good and the true. These views in turn provide a basis for their choices of which laws to impose. Giving people the freedom they are owed according to liberal theory consequently threatens that some citizens will frequently impose laws for reasons which others cannot accept, thereby disrespecting them as free and equal. Liberal theory, in short, cannot realize a plausible set of its own principles.

The solution that public reason liberals offer to this problem is for citizens to constrain their use of democratic power by what they call the *Public Justification Principle*: laws are legitimate – and so permissibly imposed – only if they are publicly justified. A law is publicly justified if and only if all reasonable citizens have sufficient reason of the right kind to accept it. Exactly what constitutes the right kind reason is a matter of debate – one to which I presently aim to contribute. Yet all public reason liberals agree that the right kind of reason must to some extent be reflectively accessible from the perspectives of reasonable citizens to plausibly solve the problem just described. If theism is objectively correct, atheists may have sufficient reason of some kind to accept laws on the basis of theistic considerations, but this reason would not be the right kind to amount to a public justification. The right kind of reason must to some extent be recognizable to atheists by their own best lights, as well as to the theists who give justifications on its basis.

For anyone who considers this idea of public justification and its motivations plausible, I offer my own problem and solution. To start appreciating this

---

1   My focus will be on statements of this idea within the Public Reason Liberalism Tradition, such as those made by Rawls (2005: 67), Larmore (2008: 148–149), Quong (2011: 2), and Gaus (2011: xv–xvi). I nevertheless anticipate my conclusions to be consequential for all philosophers who endorse a public justification requirement irrespective of whether they claim membership in this tradition, such as deliberative democrats like Gutmann and Thompson (2004: 7) and theorists of legitimacy like Applbaum (2019: 252).

problem, note that 'public justification' is ambiguous. In the Public Justification Principle, this term refers to the *state* of all reasonable people having sufficient reason of the right kind to accept a law. In the vast majority of cases, the only way that a law can achieve this state is by reasonable people depending on interpersonal justifications given to them by speakers through news reports, social media, government websites, and other lines of communication.[2] Accordingly, 'public justification' can also refer to the *act* of giving an interpersonal justification for a law that contributes to the law's achieving the state just described. A person performs this act just when her interpersonal justification gives at least one reasonable person sufficient reason of the right kind to accept the law.

But not just any justification should contribute to a law's legitimacy. Public reason liberals need a criterion for when interpersonal justifications count as public justifications. To this end, all proponents of what I call the Standard Conception of public justification assume that public justifications must be representable as arguments that the reasonable listeners to whom they are addressed would accept.[3]

This criterion would be unsatisfied if Addison appealed to the movements of a dowsing rod or crystal pendulum to justify where to drill a publicly funded well to Brianna, who considered these techniques unscientific hocus-pocus. It would be unsatisfied even if Addison explicitly argued that the drilling policy was based on sound science, Brianna rationally accepted this argument, and Addison implicitly based her decision exclusively on the dowsing rod and crystal. Brianna might have sufficient reason of some kind to accept the law in this case, but it would not be sufficient reason of the right kind for a public justification. Satisfying the criterion requires that Addison at least in principle can display the policy as sufficiently supported by an argument composed of premises which Brianna would accept and that is adequately elaborated so as to not substantively require Brianna to take Addison at her word. Addison cannot do this considering the full range of reasons on which she essentially bases the policy.

This criterion is plausible given the prevalent idea that reasonable citizens ought to respect each other as free and equal in their capacity to reason. Showing this kind of respect demands that reasonable citizens motivate each other to accept laws by using rational persuasion, as opposed to threats, deception,

---

2   Donahue (2020: 379).
3   For expressions of this criterion, see Rawls (1997: 786; 1999: 508–509), Mason (2007: 679), Gaus (2010: 184; 2011: 244–245, 247 236), Quong (2011: 141–142), Hartley and Watson (2018: 20, 84–85), and Vallier (2019: 97–98), among others.

emotional manipulation, and so forth. The paradigmatic way to rationally persuade someone without disrespectfully circumventing her capacity to reason is to offer her a valid argument with premises which she would accept. Not only does this make some notion of argument the natural place to reach for a criterion for which acts are public justifications, it is difficult to see an alternative that could play this role.

I nevertheless claim to the contrary that public justifications are not arguments.

A law's being in a state of public justification does not require every reasonable person to be able to accept the law on the basis of some adequately elaborated argument which she would accept, even in principle. I argue for the alternative criterion that a law's being in a state of public justification requires every reasonable person to have sufficient opportunity to accept the law by depending on interpersonal justifications from speakers with adequate moral character. This criterion can be satisfied even if some reasonable people cannot accept the broader considerations on which these speakers base their interpersonal justifications.

Adopting this alternative conception of public justification is the solution I propose to what I call the Dark Knowledge Problem. A person has dark knowledge that p just when she knows that p but would rationally reject the belief constituting her knowledge were she to learn particular details about its origin. More specifically, dark knowledge is knowledge which you have only because you are sufficiently in the dark about epistemically relevant aspects concerning how your belief came about. Suppose that relying on computer simulations to decide where to drill wells is sound science but that Brianna falsely considers it hocus-pocus. She nevertheless holds this belief rationally on account of responding as best she can to her available evidence.[4] Cecilia, a credentialed government hydrologist, informs Brianna that drilling in Spot-A will access water while not mentioning that her claim relies on computer simulations. In the appropriate communicative circumstances, if Brianna believes what she is told, her belief that drilling in Spot-A will access water counts as dark knowledge.

Dark knowledge creates a problem for the idea that public justifications are arguments in the following way. People who are at least moderately free to reason by their own lights naturally develop views and confront problems of great complexity. The sincere attempt to publicly justify certain laws under these circumstances frequently has the unintended and unavoidable

---

4 One might be incredulous that Brianna could hold such a belief rationally. I elaborate this example to show how this can be possible below in Section 4.

side-effect of some reasonable listeners becoming rationally persuaded to accept such laws on the basis of interpersonal justifications that produce dark knowledge. These justifications cannot be represented as arguments from premises that reasonable listeners would accept, however, and so fail to count as public justifications according to the Standard Conception. Liberal theory, in short, remains unable to realize a plausible set of its own principles in spite of the repairs to address the Fact of Reasonable Pluralism.

Because this Dark Knowledge Problem and the Fact of Reasonable Pluralism arise from the same circumstances of justice, insisting on adhering to the Standard Conception in response to the Dark Knowledge Problem conflicts with the considerations that motivated accepting the Public Justification Principle in the first place. The best way out of the problem is to abandon the Standard Conception with its assumption that public justifications are arguments. We accordingly require a new criterion for when interpersonal justifications are public justifications.

I suggest that only speakers adequately free of civic vice have standing to publicly justify laws. Not every kind of civic vice is relevant to public justification, but only those that reasonable people agree deserve exclusion. I sketch a view of such vices and their importance for regulating political deliberation. However, my primary aim is to explain the Dark Knowledge Problem and the revisions it requires for the theory of public justification. Proponents of this theory have considered civic virtue at most instrumentally beneficial for achieving a state of public justification.[5] My view is that civic virtue makes the difference between mere interpersonal justification and public justification. Because civic virtue plays this role for the most part independently of the further reasons on which those justifications are based, it is an essential part of the state of public justification itself.[6]

1   The Standard Conception of Public Justification

Before striking out into this new territory into which the Dark Knowledge Problem forces us, we need to get a better idea of the Standard Conception

---

5   See, for instance, Mason (2007: 679–681), Schwartzman (2011: 377–378), and Boettcher (2012: 161).

6   I say 'for the most part' because deliberately basing an interpersonal justification on some kinds of reasons, such as racist or sexist ones, is a way of having civic vice. Audi (1998: 162–163) presents a broadly similar view in that he considers possessing civic virtue necessary for fulfilling obligations of conduct in political discussions irrespective of the content of one's contributions. He eschews linking his ideas to the necessary conditions for public justification, however (ibid.: 158).

of public justification. This conception is fundamentally motivated by what we may call the *Circumstances of Justice Principle*: a political theory's normative principles ought to be jointly satisfiable under the best foreseeable circumstances of justice.[7] The circumstances of justice are the conditions under which reasonable people must work together to establish and maintain just political rule.[8] The best such foreseeable circumstances are those of a realistic utopia where all people are reasonable and voluntarily comply with the rational and moral standards that they set for themselves.[9] This utopia is realistic in that it presupposes the physical limitations of the actual world, such as a moderate scarcity of resources and people's inherent cognitive limitations. Failing to satisfy the Circumstances of Justice Principle accordingly means that a theory's principles are unrealizable even within a moral scenario that is better than any which we are likely to encounter.

As mentioned earlier, the Public Justification Principle is motivated by emphasizing a particular circumstance of justice called the Fact of Reasonable Pluralism.[10] Again, this refers to the fact that people who use their rational capacities under conditions of freedom inevitably form conflicting views, such as those of Christians, Atheists, Libertarians, Wiccans, Socialists, and so forth. People can moreover be *reasonable* in adopting these views even if they turn out to be false.[11] That is to say, they can accept their views as the result of reasoning according to their best lights and while sincerely wanting to cooperate on fair terms.

The Fact of Reasonable pluralism is not a basic circumstance of justice but rather is the outcome of several others. Heading the list philosophers typically offer is that the evidence bearing on which laws to accept is complex.[12] We should add, as a corollary, that the time to evaluate it is limited. It is not hard to see that reasonable pluralism results from these circumstances even in a

---

7   For expressions of this principle in the broadly Rawlsian view of Public Reason Liberalism, see Rawls (2005: xvii, 66) and Quong (2011: 37, 143). Within a broadly Gaussian view, this principle is implicit in the commitment to holding realistic expectations of people given their actual capacities for reasoning. See Gaus (2011: 235–258).

8   Rawls (1999: 109–112).

9   Quong (2011: 92).

10   See Quong (2011: 6) and Gaus (2011: xv–xvi).

11   Rawls (2005: 54–58).

12   See, for instance, Rawls (2005: 56), Quong (2011: 37), and Vallier (2019: 20). In an earlier discussion, Rawls says that morally faultless pluralism results from incomplete knowledge and the limitations of our "powers of reasoning, memory, and attention" without mentioning evidential complexity (1999: 110). Evidential complexity is nevertheless an obvious source of incomplete knowledge when coupled with the limitations that he mentions. It is therefore at least an indirect source of pluralism, one that people face in any contemporary society.

realistic utopia. Just consider the variety of opinion which we find among well-meaning experts about how to address issues like climate change, economic inequality, or pandemic response. Lastly, the emergence of reasonable pluralism requires the existence of non-authoritarian political regimes where citizens are at least moderately free in that their disagreement is not harshly suppressed.[13]

A consequence of these observations is that liberal political theories are in danger of violating the Circumstances of Justice Principle.[14] Liberal theories assume that citizens ought to live under non-authoritarian political regimes. Such theories also typically assume that citizens ought to respect each other as free and equal in their capacity to reason. Reasonable citizens living in non-authoritarian regimes naturally develop a plurality of conflicting views that motivates them to frequently impose laws for reasons that other reasonable citizens cannot accept. Yet imposing laws on others for reasons which they cannot accept disrespects them as free and equal. You would be disrespected as equally capable of reasoning, for instance, if other citizens dismissed your sincere objections and forced you to conform to laws based on religious views or empirical claims which you would not see sufficient reason to accept.

Public reason liberals argue that avoiding this result requires liberal theories assuming the principles just mentioned to also assume the Public Justification Principle. Commitment to this principle accommodates the Fact of Reasonable Pluralism because it is a commitment to a kind of rational transparency. Publicly justified laws are ones that all reasonable people can in some sense see sufficient reason to accept from their own perspectives.[15] Imposing laws only if they are in a state of public justification respects reasonable people as free and equal because it entails that all reasonable people can in principle be rationally persuaded to accept the social order that these laws create.[16] The act of publicly justifying laws is furthermore necessary for respect because of how it brings this state about. The act of giving a public justification assures reasonable citizens that they will not be imposed on in ways that are unacceptable to their perspectives at the first opportunity.[17]

Beyond this general description, political philosophers have various opinions about how to interpret the Public Justification Principle. Yet they all assume that interpersonal justifications contribute to satisfying this principle

---

13   Rawls (2005: 37).
14   Quong (2011: 142–143).
15   See Nagel (1987: 232), Waldron (1987: 149), and Vallier (2020: 1112).
16   See Waldron (1987: 128), Gaus (2011: 17), Quong (2011: 2), and Vallier (2019: 9).
17   Lister (2013: 114–115).

only if they are representable as arguments.[18] Jonathan Quong most explicitly states this assumption. As he says:

> Suppose the question is whether or not the proposition Q can be justified to Peter. Suppose the only valid justification for Q depends on premises A, B, and C. In order for Peter to be justified in believing Q, Peter must therefore also be justified in believing A, B, and C. But what if there is no way to justify A to Peter? The grounds for premise A might, for instance, be an eyewitness account of some event that Peter has no good reason to accept.[19]

According to Quong, whether Peter's belief in Q has been justified to him depends on whether he can be offered additional premises that are acceptable to his perspective to support his beliefs. As Quong continues, "In order to justify Q to Peter, we must also ensure all the premises and steps in our argument can be justified to Peter, and that may or may not be possible depending on Peter's wider epistemic situation."[20]

For other writers, the assumption that public justifications are arguments lies just below the surface. Gerald Gaus and Kevin Vallier claim that reasons contribute to public justification only if members of the public can accept them by following a sound deliberative or inferential route from their perspectives.[21] Although Gaus and Vallier do not assume that people typically reason by explicitly following arguments, it is hard to imagine what an instance of sound deliberation or inference might look like that could not in principle be represented as an argument.

Given the above general description of public justification and the textual evidence just mentioned, we can see that proponents of public justification are committed to the following *Argument Representation Principle*: a speaker's interpersonal justification contributes to satisfying the Public Justification Principle only if it is representable at least in principle as an argument for a conclusion derived via premises and inference rules which the listener would accept, where this argument is adequately free of testifier-centered premises. A couple of clarificatory remarks about this principle are in order.

---

18   See Footnote 3 for references.
19   Quong (2011: 141). The second most explicit statement is from Rawls (1999: 508), who says, "... justification is argument addressed to those who disagree with us, or to ourselves when we are of two minds .... Ideally, to justify a conception of justice to someone is to give him a proof of principles from premises we both accept ...".
20   Quong (2011: 142).
21   See Gaus (2010: 184; 2011: 244–245, 247 236) and Vallier (2019: 97–98).

First, the qualifier 'at least in principle' signifies that it is admissible for an argument representing a particular interpersonal justification to contain more information than what the speaker has explicitly communicated to the listener. It can also contain the uncommunicated reasons which the speaker could sincerely offer to justify her testimony. And if any of those reasons constitute an appeal to testimony, it can also contain the uncommunicated reasons which this further speaker could sincerely offer to justify her testimony, and so forth. The Argument Representation Principle is satisfied when these justifications taken together form an argument meeting the further strictures mentioned in the Argument Representation Principle.[22]

Second, a premise is 'testifier-centered' just when a listener lacks sufficient reason to accept it apart from the speaker's credibility. One cannot be too permissive about including such premises if public justification is to provide the kind of rational transparency which the Standard Conception assumes. Suppose that an interpersonal justification is representable as the following argument: 'I'm a credible speaker, and the law I'm proposing is a good law. So, you ought to accept it.' Although hearing this argument might sometimes render a listener epistemically justified in considering the law acceptable, it does not count as a public justification because it disrespects the listener as free and equal. Having such meagerly informative testimony as the sole basis for accepting a law circumvents a listener's capacities for rational judgment rather than engaging them.[23]

That said, there is disagreement about the extent to which an argument representing an interpersonal justification must be free of testifier-centered premises. On what we may call the Consensus View, testifier-centered premises can be included if they express claims on which a relevant group of experts are in consensus, irrespective of whether these claims are acceptable to the rest of the reasonable public.[24] This view is controversial, however, since it conflicts with the rational transparency typically assumed necessary for public justification.[25] Some philosophers accordingly opt for the more restrictive Consensus+ View, on which appeal to a testifier-centered premise p is admissible only if experts are in consensus that p, reasonable citizens are

---

22   The condition described in this paragraph is a consequence of charitably interpreting Quong's (2011: 141–142) remarks quoted above, given that he focuses on what can be justified to a person as opposed to what has been justified to that person in a particular interpersonal justification.

23   Larmore (2008: 13).

24   Proponents of this view include Badiola (2018: 426), Kappel (2021: 620, 632), and arguably Rawls (1999: 480; 2005: 224). See Badiola (2018: 426) for a defense of this latter claim.

25   See Jønch-Clausen and Kappel (2016) and Reid (2019: 496).

in consensus that they ought to defer to these experts about *p*, and reasonable citizens have access to relevant information concerning *p* and how the expert consensus was formed.[26]

Finally, on what we may call the Default View, testifier-centered premises must be completely excluded.[27] Charles Larmore, for instance, says that a rule is publicly justified only if the people on whom it is imposed "could be moved to abide by the rule … by the very reasons we ourselves have for imposing the rule."[28] Thomas Nagel similarly says public justification requires "that it must be possible to present to others the basis of your own beliefs, so that once you have done so, they have what you have, and can arrive at a judgment on the same basis."[29] The principle these philosophers express is violated when appealing to an argument containing testifier-centered premises. Listener acceptance is then based irreducibly on the credibility of some speaker. Meanwhile, the speaker's acceptance or imposition of the law is based on some other reason, given that it cannot be based on a self-directed appeal to her own credibility.

Apart from the value of transparency and rational persuasion already described, the literature suggests two more reasons for assuming the Argument Representation Principle that are worth mentioning. First, there is the need to avoid exploiting listener ignorance.[30] Speakers can induce listeners to accept laws by deliberately withholding key reasons for imposing them while knowing that listeners would find them unacceptable. The Argument Representation Principle prevents testimony from contributing to public justification in such cases. Second, there are concerns about political stability. Proponents of public justification think that if a law is imposed on reasonable people on the basis of considerations which they cannot accept, they will eventually become aware of this and reject the law.[31] This need not mean that reasonable people will actively resist the law or foment insurrection. But at the very least, they will not view the law as acceptable from their own perspectives and so cannot consider themselves respected as free and equal members of their political community in complying with it.

---

26   Reid (2019: 492) endorses this view.
27   The term 'Default View' is apt because other views divert from the typical expressions of public justification's transparency requirements and so need special argument for support.
28   Larmore (2015: 78).
29   Nagel (1987: 232).
30   See Quong (2011: 265–267), Schwartzman (2011: 386), and Hartley and Watson (2018: 45).
31   Rawls (2005: 119).

2   Dark Knowledge

This suffices to characterize the Standard Conception of public justification. We can now turn to characterizing dark knowledge. For a more precise description of the concept, let us say that a person, S, has dark knowledge that *p* if and only if the following conditions are satisfied:

(i)   S knows that *p*.
(ii)   S's current rational beliefs would defeat S's justification for believing *p* were S informed of the relevant processes by and reasons for which S acquired this belief.
(iii) No possible world in which S gains the information mentioned in Condition (ii) is nearby.

The relevant processes and reasons mentioned in Condition (ii) are all and only those that S is rationally disposed to consider relevant to epistemically evaluating how S's belief that *p* came about, holding fixed S's current rational beliefs.

This relevance requirement prevents the stock of details bearing on whether someone has dark knowledge from becoming arbitrarily expansive. When we want to epistemically evaluate how our beliefs came about, we care to learn the causal processes leading to them that we think make a difference to their truth or justification. And insofar as persons are involved in these processes, we care to learn if they acted for and can support their contributions with reasons that are similarly important. The great-great-grandmother of the testifier from whom I learned that *p* is in some sense involved in the process by which I acquired the belief that *p*; after all, if the great-great grandmother did not exist, I could not have gained this belief by interacting with the same testifier. But unless it is reasonable for me to consider the identity of my testifier's great-great-grandmother relevant to the truth or justification of my belief, this detail has no bearing on whether I have dark knowledge.

This is just to say what dark knowledge is not. For an illustration of what dark knowledge is, consider an example that I call *Misleading Education*. Suppose you have acquired a simplified view of science as a result of responding as best you can to the evidence available to you. I mean this in the following sense: First, you have gained this view by responding appropriately to the evidence that you have been exposed to, such as through your education and depictions of science in general culture. Second, it is not reasonable to expect you to develop a more complex view given your interests. And third, the evidence you would find if you spent a bit more time investigating your beliefs would not easily undermine your view. This simplified view of science is accordingly one that you hold rationally. As a result, you are reasonably disposed to reject some

investigatory practice P as unreliable, even though P is reliable and supported by the best current science.

One day, a reliable authority who seems credible tells you a truth: Fact Y. As it happens, you cannot easily interact with this authority or any other who could give you credible testimony about Fact Y. But if you did, the authority would truthfully inform you that she believes Fact Y because it was discovered by relying on practice P and that there are no alternative grounds on which to believe this fact. You are nevertheless ignorant of these details and so form the belief in Fact Y based on the authority's assurance.

Your belief in Fact Y is dark knowledge. You know Fact Y through the authority's testimony. Because you are reasonably disposed to consider practice P unreliable, processes involving it are relevant to the epistemic evaluation of your beliefs. Your current rational beliefs would defeat your belief in Fact Y were you informed that you acquired it through an application of this practice. Yet since you cannot easily interact with any authority who can reveal the source of your belief, no possible world is nearby in which you gain this information.

At this point, there are two concerns one might have about dark knowledge, one conceptual and the other practical. The conceptual concern is that dark knowledge is incoherent because it presupposes an impossible combination of knowledge, epistemic justification, and defeaters. The practical concern is that dark knowledge is too rare to be important. I will dispel both concerns, starting with the conceptual one.

Theories of epistemic justification are either internalist or externalist. Epistemic externalism is the view that beliefs are justified when they are the outcome of a sufficiently reliable process, irrespective of whether the person who forms them could become reflectively aware that the process is reliable.[32] Dark knowledge is compatible with externalism. Consider again Misleading Education. Fact Y was discovered by relying on practice P, and P is reliable by stipulation. Furthermore, you learn about Fact Y from a reliable authority who is credible from your perspective. Depending on testimony from credible and reliable authorities who base what they say on reliable practices is itself a reliable way of forming beliefs. Given how often we depend on authorities to form our beliefs, very few of our beliefs would be justified by the externalist's lights if this process were insufficient to externally justify your belief.[33]

---

32   Bonjour (1980: 56).

33   The most significant threat to your belief's externalist justification in this case is that you are disposed to prefer testifiers who do not base their claims on procedure P. Yet even if you have this disposition, it cannot be plausibly construed as a constitutive part of the way that you formed your belief and so does not undermine the example.

Epistemic internalism is the view that beliefs are justified when they are adequately supported by features about which the believer can easily become reflectively aware.[34] Dark knowledge is also compatible with internalism. On standard internalist views, that a seemingly credible testifier has told you that *p* provides sufficient, reflectively accessible support for the belief that *p*. Epistemologists disagree about whether credible testimony is a fundamental source of justification or one that reduces to others, like sense perception, memory, and so forth. This dispute is nevertheless irrelevant for our concerns because epistemologists agree that credible testimony is at least a source of epistemic justification. In Misleading Education, a credible testifier has told you Fact Y, thereby giving you sufficient support for your belief in Fact Y that is easily accessible upon reflection. Again, given how often we depend on authorities to form our beliefs, very few of our beliefs would be justified if these features of your situation were insufficient by the internalist's lights.

Since these are all the views of epistemic justification on offer, there is no good reason to think that your belief in Fact Y is unjustified. Fact Y is also by stipulation true. This does not suffice to show that dark knowledge is conceptually coherent, however, because it is a commonplace of epistemology that true, justified beliefs sometimes fail to be knowledge.[35] The only plausible reason why your belief in Fact Y could be true, justified, but not knowledge in this case is that some defeasibility condition for knowledge is violated. There seemingly is a defeater for your belief in Fact Y, namely that practice P is an essential source of support for Fact Y. Although your belief in Fact Y is justified now, it would become unjustified were you to learn that practice P is an essential source of support for Fact Y.

However, the existence of this kind of detail is no threat to the conceptual coherence of dark knowledge. Given any of our beliefs, there is almost always some feature existing somewhere – a stray bit of testimony, a fact about our situation, and so forth – that could similarly act as a defeater were we to become aware of it in isolation.[36] Avoiding skepticism requires that such truths must be easily encountered before they can defeat knowledge in this way. In Misleading Education, the truths that would make your belief in Fact Y unjustified were you to learn them are by stipulation not easily encountered and so a defeasibility condition for knowledge is not violated. Based on these considerations, we should conclude that dark knowledge is a coherent concept.

---

34   Goldman (1999: 272).
35   Gettier (1963).
36   Feldman (2003: 34–36).

This brings us to our second, practical concern. Even if dark knowledge exists, the concept would be insignificant if it were rarely involved with reasonable judgments of which laws to accept. This concern is also unfounded. We should expect political deliberation about complex issues to frequently depend on dark knowledge, even in a realistic utopia where all people are reasonable and voluntarily comply with the rational and moral standards they set for themselves.

This is because of the practical demands of communication. The evidence bearing on which laws we ought to accept is complex and the time any individual has to evaluate it is limited. Forming reasonable judgments about which laws to accept consequently requires a division of epistemic labor in which we depend on testimony from speakers whom we judge credible. These speakers may be individual persons (a president, chancellor, scientific expert, etc.) or institutions (a department of government, a policy research center, a corporation, etc.).

Now, consider what these relations of dependence involve. It is too practically demanding for speakers to make explicit all the considerations on which they base their testimony.[37] It is similarly too practically demanding for listeners to make explicit all the considerations that would lead them to reject testimony if they learned speakers relied on them. Speakers and listeners of course can sometimes communicate with each other to make these details explicit. But in many contexts – such as when speakers and listeners are too numerous or too disconnected to give each other this kind of attention – communication is a one-way street. Listeners have no information about why speakers think what they say is correct beyond what speakers have time to publicize, and speakers must guess at what considerations listeners would treat as grounds for rejecting their testimony.

Interpersonal justifications for laws easily produce dark knowledge in such contexts. All that is required is a pairing of speaker and listener such that some of the considerations on which the speaker implicitly bases her testimony are ones the listener would reject. If the listener and speaker are sufficiently disconnected from each other and the speaker is a reliable testifier, then it can be very difficult for the listener to become aware of these considerations. The speaker's interpersonal justification for the law can then create dark knowledge on the listener's part.

Such contexts moreover frequently arise considering the size of most societies and the complexity of the reasons bearing on which laws to impose. Responsibly crafting laws requires integrating numerous empirical and moral

---

37   See Millgram (2015: 28, 30) and Medvecky (2020: 82).

considerations. This holds for laws about constitutional essentials, like the government's structural features and basic rights and liberties. It also holds for laws about topics of general concern, like environmental regulations, public health interventions, economic policies, and so forth. Regulations to curb tobacco consumption, for instance, might draw on research from economics, sociology, epidemiology, neurobiology, chemistry, and other disciplines and demand weighing values about autonomy, public health, the environment, and other topics.[38]

Combining many expert views and publicizing the result creates relations of testimonial dependence like those just described. If we assume that an interpersonal justification of a law about a complex issue has only a very small chance of producing dark knowledge for any given person – a tenth of a percent, for instance – then it becomes highly likely that dark knowledge will be produced in a significant minority of people when that justification is publicized to tens and hundreds of millions of individuals. We should consequently expect basing judgments about which laws to accept on dark knowledge to frequently occur, especially among large aggregates of reasonable people.

3   The Dark Knowledge Problem

We are now in a position to incorporate the results from the previous sections into the Dark Knowledge Problem. To see the problem, consider a schematic situation involving a listener and credible speaker who are both sincerely committed to the ideal of public justification and to respecting reasonable people as free and equal. The speaker is accordingly disposed to refrain from imposing a law if she believes that it is unsupported by considerations which reasonable citizens can accept, and the listener is disposed to reject a law if she believes the same.

Suppose the speaker aims to publicly justify a law by addressing an argument consisting of premises $P_1, \ldots, P_n$ to an audience that includes the listener. We can assume that the argument is sound, that the speaker has been reliable in giving her testimony, and that receiving the argument rationally persuades the listener to accept the law. Because the speaker is both credible and reliable, the listener knows each of these premises. Suppose, however, that the listener essentially relies on one of the premises P_i as an instance of dark knowledge. By 'essentially relies,' I mean that considerations reflectively accessible to the

---

38   National Cancer Institute (2007: 26).

listener apart from the premise leave her insufficient reason to accept the law, and there are moreover no considerations aside from the ones which the speaker presently depends on that she can sincerely communicate to the listener to replace this premise or provide it alternative support.

If we assume the general kind of communicative contexts described in the previous section, there are various ways in which this schema could be instantiated. First and most straightforwardly, P_i might be a scientific claim like Fact Y that the speaker judges too confusing or time-consuming to defend but that turns out to be supported by investigative practices which the listener would reject.

Second, this schema can easily arise because the immediate objects of scientific consensus typically require interpretation to become intelligible to listeners. The immediate objects of scientific consensus are often complex claims about probabilities that laypersons easily misunderstand.[39] For example, empirical research suggests that laypersons often assume that events described as 'Very Likely' (intended by speakers to correspond to a probability above 90%) have a probability close to 70% or even lower.[40] Expressions of scientific uncertainty can also be easily taken for lack of adequate proof.[41] Speakers might try to avoid such problems by giving reasonable citizens more access to probability claims and information about how to interpret them. This is no solution, however, because many citizens are simply too busy to use such resources.[42]

Under these circumstances, speakers aiming to publicly justify laws can reasonably do so by framing scientific results as more certain than they are. P_i can be one of these simplified claims. Reasonable listeners can know P_i just as we know simplified claims such as that smoking causes cancer, that vaccines are safe, that carbon emissions cause global warming, and so forth. At the same time, it is plausible that many reasonable people would reject simplified claims like these as admissible interpretations of the results of science were they to learn the more complicated probabilistic statements that are the true objects of consensus. They would accordingly be disposed to reject the laws which they presently accept insofar as the simplified claims motivate their acceptance.

And third, reasonable listeners can believe P_i through so-called elite cue-taking. It is reasonable for citizens to sort through the complex information

---

39   Kappel (2021: 621).
40   Budescu et al. (2014).
41   Oreskes (2004).
42   Donahue (2020: 383).

confronting them by accepting policies based on cues – talking points, slogans, and other simplified messages – from trusted political leaders, media personalities, and other public figures.[43] Empirical research suggests that this sometimes results in citizens adopting policies against their earlier professed views, for instance by opposing government benefits which they supported before learning a favored politician's stance on them.[44] If the cue is given by a responsible and well-meaning public figure, this change can be for the better. It can cause citizens to adopt policies and acquire knowledge which they would reject if they compared the public figure's underlying reasons to the reasons supporting their own initial views. The argument featuring P_i could itself be such a beneficial cue, or P_i could be adopted as a response to such a cue.[45]

We can by now see a pattern running through these examples. A speaker can sincerely believe she is reasoning in ways which reasonable listeners would accept, simplify her thoughts in a sincere and even beneficial effort to communicate, and yet have it turn out that listeners would reject her claims if they learned the more complex, underlying considerations. This pattern is furthermore not avoided by adopting a Consensus or Consensus+ View. The second and third examples show that the objects of a scientific consensus can themselves form the more complex considerations that speakers must simplify in order to communicate. In this way, speakers can cause listeners to accept laws through inadvertently imparting dark knowledge. We are accordingly left with the question: Can a law be publicly justified in the kind of situation described in these examples?

The strongest reason in favor of a negative answer is that the Argument Representation Principle is violated. The listener essentially relies on premise P_i as an instance of dark knowledge. There are accordingly no alternative considerations which the speaker could sincerely appeal to in persuading the listener to accept P_i, and the listener would reject this premise and the law it supports were she to learn the considerations on which it is based. The only source of the listener's sufficient, reflectively acceptable reason to

---

43   See Downs (1957: 233) and Lupia and McCubbins (1998: 4–5).
44   Lenz (2012: 184–185).
45   There are yet more ways to instantiate the schematic situation thanks to a cousin concept of dark knowledge that we may call *dark moral belief*. It can faultlessly occur that a listener accepts a law on the basis of moral beliefs she would reject were she to learn the normative assumptions of some speaker whom she is presently relying on to have that belief. The Dark Knowledge Problem therefore cannot be dispelled merely by criticizing my use of examples with empirical content. Properly establishing the possibility of dark moral belief requires entering into epistemological and metaethical considerations that would detract from the brevity and intuitive force of the present work, however, so I will not support this claim further here.

accept premise P_i is that the speaker is credible. Consequently, there is no interpersonal justification of the law that can be represented as an argument adequately free of testifier-centered premises that the listener would accept. If we accept the Argument Representation Principle, we must reject that the law is publicly justified.

The Dark Knowledge Problem arises, however, in that the Argument Representation Principle conflicts with other, more fundamental principles of public justification. To start, note that the strongest considerations motivating the Argument Representation Principle do not apply. There is no restriction on the number of premises P_1, ... , P_n in the speaker's argument. The only stipulation is that at least one premise essential to this argument is an item of dark knowledge for the listener. This argument consequently provides a basic amount of transparency in that the law can be explained to the listener for the most part in terms which she can accept. What the speaker says is not equivalent to 'Trust me, you ought to accept the law.'

Next, consider the worry about exploiting listener ignorance. It is true that the listener is ignorant about features of her epistemic situation which she plausibly cares about, namely that the reasons on which the law is based are ones which she would not reasonably accept. No one is exploiting her ignorance, however. The speaker has made a sincere attempt at and moreover believes that she has succeeded in justifying the law on the basis of considerations that all reasonable people would accept. She respects the listener as free and equal in that she would refrain from imposing the law if she knew that the listener would reject the reasons on which her testimony is based. The speaker's communicative situation ensures that she cannot easily gain this knowledge, however.

Finally, there is the issue of stability. The listener's acceptance of the law is stable because she knows each premise P_1, ... , P_n. This means that none of her beliefs in these premises are defeated. There is consequently no close possible world in which she learns the full range of reasons on which the speaker's testimony is based and rejects the law. The listener's reflectively accessible reason to accept the law is moreover based on the speaker's credibility and the plausibility of her testimony's content. She is not being coerced into accepting the law. Her acceptance is therefore stable for the right reasons.

The last point to make is that liberal political theories assuming the Argument Representation Principle violate the Circumstances of Justice Principle. We saw in the previous section that we should expect justifications to frequently depend on dark knowledge when speakers communicate to large aggregates of reasonable people about political issues where the evidence bearing on them is complex and the time to evaluate it is limited. Complexity

of evidence and limited time were the same circumstances that led to the Fact of Reasonable Pluralism. Furthermore, like reasonable pluralism, we should expect this dependence to frequently occur even in a realistic utopia where reasonable people are at least moderately free to reason by their own best lights. However, justifications that depend on dark knowledge fail to satisfy the Argument Representation Principle. We should therefore expect this principle to frequently be unsatisfied even in a realistic utopia. Liberal political theories assuming the Argument Representation Principle accordingly violate the Circumstances of Justice Principle.

This Dark Knowledge Problem gives proponents of public justification a choice: they must either reject the Circumstances of Justice Principle, reject requiring conditions of at least moderate freedom, or reject the Argument Representation Principle. Taking the first two options means rejecting principles that were essential to motivating the Public Justification Principle in the first place. The best response is accordingly to reject the Argument Representation Principle. This means rejecting the Standard Conception's assumption that public justifications are representable at least in principle as arguments that listeners would accept.

4   Rejecting the Argument Representation Principle

Let us assume, therefore, that the Argument Representation Principle ought to be rejected. A concern one might have at this point is that the consequences of doing so are insignificant. Perhaps some weakened version of the Argument Representation Principle avoids the Dark Knowledge Problem. If this weakened version is moreover compatible with the theory of public justification as currently formulated, a trivial response to the problem is to accept this principle without substantially revising the theory. Yet this concern is unfounded. I argue in this section that the Dark Knowledge Problem is not this easily solved.

An initial point to note is that the Argument Representation Principle is in various respects already a weak principle. As mentioned earlier, it requires interpersonal justifications to be merely hypothetically representable as arguments. Their content accordingly does not depend on what speakers have time to represent or what listeners have time to cognize. Although these arguments must be adequately free of testifier-centered premises, the Argument Representation Principle is already formulated to allow their inclusion up to the limit of what would otherwise require revising the theory of public justification. For these reasons, weakening the requirement that

public justifications must be representable as arguments is implausible. A trivial response must be based instead on weakening the requirement that these arguments are ones that listeners would accept.

There is seemingly precedent for this kind of revision. Some philosophers formulate the Public Justification Principle using weaker modals.[46] Considering John Rawls is instructive here. If the Argument Representation Principle were recast with the same modals he often uses to express the Public Justification Principle, it would read: 'Public justifications must be representable as arguments that listeners *may reasonably be expected to accept.*'[47] This principle is weaker than my statement of the Argument Representation Principle because a person may reasonably be expected to accept that p even though she would not accept that p. What modal this formulation expresses is ambiguous. This nevertheless has the virtue of allowing us to consider other weakenings of the principle. The upshot for our purposes is that the Dark Knowledge Problem remains unsolved on any plausible resolution of this ambiguity.

Specifically, what a person may reasonably be expected to accept can be read either normatively or predictively.[48] On the normative reading, an argument which a person may reasonably be expected to accept is an argument which the person would accept if she were idealized to have the traits of reasonableness, such as by being disposed to cooperate on fair terms, competently weigh evidence and draw inferences, and so forth.[49]

This disambiguation cannot solve the Dark Knowledge Problem. This is because the problem arises even when persons are already normatively idealized as reasonable. We can illustrate this by using the example from the beginning of the article of Brianna and Cecilia, the well-drilling, government hydrologist who implicitly relies on computer simulations. With some plausible elaborations, this example becomes an instance of the schematic situation involving Dark Knowledge mentioned in the previous section.[50]

---

46 To give a couple of additional examples: Jonathan Quong remarks that public justifications "appeal to what reasonable people could accept" (2011: 10). Amy Gutmann and Dennis Thomson state that citizens and representatives in deliberative democracies should appeal to "reasons that should be accepted by free and equal persons seeking fair terms of cooperation" (2004: 3). I thank an anonymous referee for raising this issue.

47 I take this phrasing from what Rawls in *Political Liberalism* calls the liberal principle of legitimacy (2005: 137, 217).

48 In making this distinction, I roughly follow Leland and van Wietmarschen (2012: 725–726).

49 Rawls (2005: 50, 55).

50 One can imagine still more cases that illustrate the same point by considering the various ways in which the Dark Knowledge Problem or something structurally similar can arise that were mentioned above.

To this end, suppose that Brianna and Cecilia are both disposed to cooperate on fair terms, have all the powers of thought to count as reasonable, and cannot reciprocally interact with each other easily (they must exchange emails though a government website). Suppose in addition that Brianna was taught in school that science requires empirical experimentation. However, given her projects and interests, namely running a bed and breakfast in a rural area, she has no good reason to update this belief. This leads her to believe that empirical claims made on the basis of computer simulations are not scientific. Furthermore, this belief is not easily undermined in that Brianna would find various ways to defend it if prompted to conduct some research. She would appeal to prominent philosophers of science who say that computer simulations do not causally interact with physical systems and so are not measurements of these systems.[51] She would point to textbooks that seemingly state that computer simulations provide information about abstract models, not the physical world.[52] And so forth. Brianna's stance is not that computer simulations have no place in science (she might believe that they are good for initially formulating hypotheses). Instead, she believes they are inadequate to establish specific empirical claims, especially when the stakes are sufficiently high. Given these details, this is a belief that Brianna holds rationally because she is responding as best she can to her available evidence.

Next, suppose the stakes are sufficiently high from Brianna's perspective. This is because Cecilia's decision about well-drilling impacts Brianna's interests. The simulations indicate that the best place to drill is on Brianna's land, which risks negatively impacting her income and enjoyment of her property. Brianna nevertheless agrees to have her property drilled because she trusts Cecilia's claims that doing so is supported by the best science and will adequately promote the public good. We may suppose that Brianna's trust is also rationally placed. Suppose that Brianna had similar interactions with government scientists in the past, and after scrutinizing their claims further, she found nothing objectionable. She consequently has no reason to believe

---

51   See Giere (2009: 60–61).

52   For instance, Nigel Gilbert and Klaus G. Troitzsch remark in their textbook *Simulation for the Social Scientist*: "while simulation has similarities with experimentation, it is not the same. The major difference is that while in an experiment one is controlling the actual object of interest … in a simulation one is experimenting with a model rather than the phenomenon itself" (2005: 14).

that this case is any different, even though it is: the scientist before her is *now* basing her claims on computer simulations.[53]

Given the information available to her, it is reasonable for Brianna to accept that her land ought to be drilled. At the same time, she rationally believes that computer simulations provide inadequate support for drilling on her property as a result of responding as best she can to her available evidence. Based on these details, it would be irrational for her to believe Cecilia's claims if she were informed of the considerations on which they are based. Brianna therefore could not reasonably accept these claims if she were so informed. Since Brianna is already acting reasonably, idealizing her as reasonable would not change this outcome.

At this point, one might be tempted to normatively idealize Brianna beyond merely being reasonable. For instance, one might consider what Brianna would believe if she were less devoted to her business, if she were educated differently, if the evidence bearing on Cecilia's claims were less complex, and so forth. This temptation should be resisted. Proponents of public justification undermine the motivation for their view if they use such idealizations to address the Dark Knowledge Problem. Limitations of people's powers of reasoning and facts of their circumstance that they are unable to easily control are the roots of reasonable pluralism. If the Dark Knowledge Problem can be removed by idealizing these features, so can the Fact of Reasonable Pluralism itself.

This leaves us with the predictive reading of what a person may reasonably be expected to accept. Under this interpretation, an argument a person may reasonably be expected to accept is an argument which the person is reasonably thought likely to accept from the perspective of whoever offers the interpersonal justification.[54]

---

53   The reader might be concerned that Cecilia makes no offer to provide evidence for her claims. If this is the case, assume that she makes this offer but that Brianna has insufficient reason to accept it given her track record of positive interactions with scientists and the amount of time researching the issue would take from managing her business. We could also suppose that even if she did look at the evidence, she would have to expend a significant amount of effort to determine that Cecilia's claims fundamentally rest on computer simulations in a way that she would reject. These additional assumptions illustrate that it can be difficult for persons to interpret evidence even if it is available, thereby ensuring the appropriate conditions for dark knowledge.

54   Rawls occasionally writes in a way that suggests this reading. Describing the criterion of reciprocity in *Political Liberalism*, he says, "our exercise of political power is proper only when we sincerely believe that the reasons we offer for our political action may reasonably be accepted by other citizens as a justification of those actions" (2005: xliv). This is not strong, textual evidence in favor of this reading's adoption, however. As I describe in the next paragraph, it is at odds with Rawls' own descriptions of public justification's

The predictive reading is also unable to realize the concern animating this section. Even if this reading can solve the Dark Knowledge Problem, it cannot do so *trivially*. This is because it is incompatible with the structure of public justification as described by the view's proponents. John Rawls says that the content of public justifications are drawn from a political conception of justice, which "provides a publicly recognized point of view from which all citizens can examine before one another whether their political and social institutions are just."[55] This political conception accordingly constitutes a "shared basis" for public justification.[56] Most proponents of public justification have followed Rawls in this regard.[57] Others argue that public justifications can be based on a convergence of reasons that have different but incompatible grounds.[58] In this structure, all persons have reasons that provide sufficient support from their perspectives for whatever law is imposed, although they cannot reciprocally accept each other's reasons as having justificatory force.

Compare this to the structure which public justifications can have when based on dark knowledge. Arguments composed of reasons which we think that others are likely to accept (but which they in fact would not) obviously fail to proceed from a shared basis. When the listeners to whom our arguments are addressed moreover lack sufficient, alternative support for the same conclusions, the structural requirements of the convergence view are also not met. Yet depending on dark knowledge can result in justifications with both features: the listeners involved in the Dark Knowledge Problem would not accept the reasons comprising the arguments addressed to them and also lack sufficient, alternative support for the same conclusions. The predictive reading consequently is able to solve the Dark Knowledge Problem only though presupposing non-trivial revisions to the theory of public justification.

Finally, the predictive reading is unable to offer a plausible solution to the Dark Knowledge Problem in the first place because the aspects of interpersonal justifications it emphasizes are morally irrelevant to public justification. A person is not plausibly respected as free and equal just because the speakers imposing the law on her merely think that she is likely to accept their

---

structure. This phrasing is also inconsistent with how he elsewhere formulates the public justification requirement. Much later in *Political Liberalism* Rawls endorses an actual acceptance requirement by saying that in giving public justifications "we are to appeal only to presently accepted general beliefs and forms of reasoning found in common sense" (2005: 224).

55   (2005: 9).
56   (2005: 115).
57   See, for instance, Quong (2011: 265) and Hartley and Watson (2018: 42).
58   See Gaus and Vallier (2009: 58).

arguments for doing so. This is because they can have this thought while also having the kinds of attitudes which the Argument Representation Principle is meant to exclude, such as wanting her to accept the law for reasons that are merely strategic or exploitative. Adding that those imposing the law *reasonably* think that this person is likely to accept their arguments does not change the case, assuming that this means their thinking this conforms to the merely epistemic or practical standards of reasonableness. What makes a difference to public justification is more plausibly some aspect of how the speakers are morally motivated or disposed in thinking that their arguments are likely to be accepted. But adding these aspects to the principle requires transcending the confines of a merely predictive reading, which is explicated in terms of epistemic rationality alone.[59]

Overall, no weakening of the Argument Representation Principle plausibly solves the Dark Knowledge Problem. Embracing this result opens an explanatory gap for theories of public justification. Public justifications arising from dark knowledge must depend on *something*; there would be no distinction between public justifications and mere interpersonal justifications otherwise. Whatever this factor is, it must be compatible with the general theory of public justification while also serving as an adequate replacement for the Argument Representation Principle. In the next section, I sketch my view about what this missing factor could be.

5   Civic Vice and Dark Knowledge

There are limits to where we can locate the missing factor. An initial thought is that it is a property of the justificatory content or reasons speakers depend on that is neither reflectively accessible nor acceptable to listeners. The most plausible view along this line is that interpersonal justifications are public justifications only if they are based on considerations that are true. Proponents of public justification should reject this view, however, because they consider allowing truth to play such a role to be incompatible with establishing a

---

59   An alternative strategy to defend the predictive reading is to observe that as long as making faultless errors about what reasons others would accept is possible, the best we can ever do is impose laws for reasons which we merely think that others are likely to accept. Yet this is no route to a trivial solution. We still need an explanation for why committing a faultless error like this fulfills – rather than excusably violates – the Public Justification Principle. This explanation for similar reasons must also depend on some aspect of agents that is not merely epistemic.

respectful political community. As Rawls says, "Holding a political conception as true, and for that reason alone the one suitable basis of public reason, is exclusive, even sectarian, and so likely to foster political division."[60] Elevating some interpersonal justifications to public justifications merely because they are true leads to the same result.

One might alternatively think that the Dark Knowledge Problem shows that only factors related to reflectively accessible justificatory content matters for public justification. Most plausibly, one could think that speakers contribute to public justification whenever their testimony provides listeners sufficient reflectively accessible reason to accept a law. Yet turning public justification into a merely internalist notion makes public justification too cheap. To publicly justify a law, all that needs to be done is ensure that it appears acceptable to reasonable people. Public justification could then come about through the kinds of exploitative behavior which the Argument Representation Principle served to exclude.

Overall, the missing factor cannot be a property of either the external or internal *content* of interpersonal justifications. I suggest instead that it more plausibly has to do with the moral *character* of testifiers. My proposal is that the relevant factor is whether speakers possess a character that adequately conforms with a public conception of civic virtue. In particular, a speaker's interpersonal justification is a public justification only if the speaker is free of substantial civic vice.

Civic virtue consists in the set of dispositions which persons can have that are beneficial to political community.[61] Civic vice by contrast involves dispositions that are detrimental to political community. A *public conception of civic virtue* has two aspects. The first is a package of civic virtues and vices relevant to the ideal of a political community where everyone respects each other as free and equal. The second aspect is a way of thinking about these virtues and vices that all reasonable people would accept. Virtues in this package plausibly include sincerity, generosity, civility, and so forth. Vices in this package include ones which all reasonable people can agree that citizens ought to lack, such as insincerity, stinginess, incivility, and so forth.

There is good reason to target substantial civic vice of this kind as uniquely relevant to public justification. Civic vice more plausibly makes a difference to public justification than mere epistemic vice because epistemic vice is

---

60   Rawls (2005: 129).
61   See Audi (1998: 149) and Brennan (2012: 315–316).

ultimately intelligible in terms of failure to track the truth.[62] Mere epistemic vice is consequently difficult to incorporate into a conception of public justification without amounting to a bare appeal to truth. And civic vice more plausibly makes a difference to public justification than civic virtue because requiring civic virtue would be unreasonably demanding. Possessing substantial civic virtue is supererogatory. Possessing substantial civic vice, by contrast, is a condition which all citizens must avoid.

Civic vice must be substantive in order to undermine one's standing to publicly justify. Minor civic vice – like the disposition to mix one's recyclables with general waste – is surely compatible with public justification. Substantial civic vices by contrast include the disposition to deceive, to be flagrantly epistemically negligent by public standards, to violate role-based responsibilities of public concern, to be racist, sexist, homophobic and, in general, undermine the possibility of political community where all persons respect each other as free and equal. Speakers with these defects may epistemically justify accepting laws with their testimony, but their words cannot publicly justify. Speakers without this defect who give the same testimony can contribute to achieving a state of public justification provided that their testimony gives listeners reflectively accessible reason to accept the law.

There is a final aspect of my view worth emphasizing. For civic vice to replace the Argument Representation Principle, whether a speaker is free of substantial civic vice must be a condition that is sometimes epistemically external with respect to the perspectives of reasonable people. We can often tell when speakers are substantially vicious, and their testimony fails to publicly justify in such cases. But we can also lack reflectively accessible reason to think that speakers are vicious when they in fact are, such as the kind of exploitative speakers whose justifications were excluded by the Argument Representation Principle. Testimony fails to contribute to public justification in such cases, even though reasonable people may believe otherwise.

The conception of public justification I have just sketched is morally plausible. Discovering that political leaders are substantially vicious typically undermines their standing to perform official functions. For instance, in July 2019 a trove of text messages sent between Puerto Rico's governor, Ricky Rossello, and his cabinet were publicized. They showed Rossello openly making comments that were misogynistic, homophobic, and that mocked

---

62 Some epistemic vices are civic vices and so matter for public justification. For instance, Meyer et al. (2021) argue that epistemic vices like sloppiness and obstinacy predict acceptance of misinformation. The view I am dismissing is that all and only epistemic vices matter for public justification.

victims of hurricane Maria. The messages thereafter became the focal point of unprecedented bipartisan protests that forced Rossello's resignation.[63]

The repugnance felt toward civically vicious persons as office holders extends to vicious persons as contributors to political deliberation. Imagine hearing someone like Rossello advocating for a law by arguing how it would make your community a better place. Even if you had no reason to believe that you were being told anything false, the situation feels repugnant. It would be natural to respond: 'Who are you to suggest what we should do?' Until you can accept the law on the basis of a justification from someone of better character, you can appropriately feel disrespected as a free and equal member of your political community insofar as you have to depend on such a person to form your judgments about the law.

This result is moreover theoretically sound. It can be explained by synthesizing two views about the context in which public justifications occur and why they are important. The first view – proposed by Christie Hartley and Lori Watson – is that public justifications are "addressed to persons in their role as participants in a shared project." This is the project of "living on terms of mutual respect with one another as free and equal citizens."[64] The second view – advanced by several authors – is that public justifications are important because they realize and affirm a morally important relationship between persons.[65] These authors identify this relationship as civic friendship.

I propose that civic vice plays the role that I am suggesting as a result of combining these two views in the following way: public justifications realize and affirm a morally important relationship between persons. These persons are unlike friends, however, in that they consider themselves to be involved in a shared project. Yet they also bear relationships to each other that are unlike those in Hartley and Watson's view because they are not involved in this project as mere co-participants. Instead, they see themselves more specifically as partners jointly committed to the project of living on terms of mutual respect with one another as free and equal.[66]

There is good reason to combine these views independently of our interest in addressing the Dark Knowledge Problem. A central motivation which Hartley and Watson offer for their view is that its adoption provides better political stability than the convergence view described in the previous section, according to which laws can be publicly justified for disparate reasons that

---

63   Puerto Rico Governor Resigns After Mass Protests (25 July 2019).
64   Hartley and Watson (2018: 42–43).
65   See Rawls (1997: 771, 786), Ebels-Duggan (2010: 55–56), Lister (2013), and Leland and van
      Wietmarschen (2017: 157).
66   The reader should not have in mind here romantic partners – think instead of teammates.

are not reciprocally acceptable between persons.[67] One could object, however, that persons are no more likely to be stably committed to the project of living on terms of mutual respect than to affirming the disparate reasons for which they consider laws acceptable.[68] The combined view I propose avoids this objection. We feel additional commitments toward partners as opposed to mere co-participants, and this in turn increases the stability of our commitment to the projects which we pursue with them. Even if one's valuation of the project wavers, one has additional reason to remain committed to it simply to not let one's partners down and to live up to their expectations.

The shared project view also supports the relationship view, namely by providing an explanation of why the relationship that public justification realizes and affirms is valuable, which the view's proponents acknowledge they lack.[69] Some relationships are valued partly on account of the project for which they are formed to pursue. For instance, teammates value their relationships with one another in part because they value the sport they play. Persons bearing the relationship to each other that public justification realizes and affirms are similar. An essential reason why they value these relationships is because they non-instrumentally value the project of living on terms of mutual respect as free and equal. Because the relationship relevant to the combined view is directed and structured by this project, I prefer not to think of it as typically undirected relationship civic friendship. Instead, I consider it to be the relationship of liberal comradeship.

The resources of this combined view are adequate to address the Dark Knowledge Problem. On this view, liberal comradeship is the relation which persons bear to one another as partners committed to the project of living on terms of mutual respect as free and equal. Living on terms of mutual respect involves settling the question of what these terms are and how they should be applied by exchanging public justifications, which must at least aim to appeal to considerations which all reasonable persons would accept. A necessary condition for public justifications to be distinct from mere interpersonal justifications, however, is that they are made by a liberal comrade. The justificatory situation otherwise would be indistinguishable from a mere

---

67   Hartley and Watson (2018: 49–53).
68   Vallier (2019: 188 fn. 32).
69   See Leland and van Wietmarschen (2017: 165). Lister (2013: 115) seemingly comes closest to endorsing something like the combined view, but he is not explicit that his relationship view requires persons to conceive of themselves as partners in a shared project. He also does not draw the same conclusions about the role that civic virtue plays in these relationships that I do here.

modus vivendi, a political arrangement where the law gains support through a chance congruence of interests. As the objection just mentioned to Hartley and Watson's view suggests, this chance congruence of interests can occur over projects as well as particular reasons. Even if a non-comrade appeals to the same reasons when offering a justification as would a liberal comrade, the modal profile of your relationship is different: someone who is your comrade would not insist on acting on you for reasons which you would reject if she found that doing so were adequately to her advantage or that you could not stop her from doing so.[70]

As a result, someone who purports to give a public justification purports to address others in the capacity of a liberal comrade. Yet being a liberal comrade presupposes a commitment to the shared project of living on terms of mutual respect as free and equal. Having substantial civic vice is incompatible with this commitment because someone who is vicious in this sense is disposed to treat persons in ways that disrespect them as free and equal. This has a significant impact on the relations which they can bear to other persons, even ones to whom they are locally respectful. Imagine playing basketball with someone who deliberately passes the ball to the opposing team at crucial moments, trips players on her own side, and is generally disposed to act in ways that undermine the project that you value. Even if this person never treated you poorly by tripping you or failing to pass you the ball, she is not just a bad teammate; she is no teammate at all. Persons with substantial civic vice are likewise not just bad liberal comrades; they are no comrades at all. Any contribution to political discussion that they offer as public justifications are disrespectfully made under the false pretense that they bear a relationship to whomever they address that they do not. Consequently, their interpersonal justifications are not public justifications.

This conclusion gains further support through G. A. Cohen's observation that the moral force of justification is speaker-relative.[71] To use a version of his example, suppose that a kidnapper reasons with a parent, 'You ought to pay me a ransom. Otherwise, your child will die an easily preventable death.' Although the kidnapper's interpersonal justification might epistemically justify the conclusion for the parent, it does not morally justify it, even though an argument with the same propositional content would provide moral justification if said by an innocent friend. In further analysis, Johann Frick argues that this effect

70   Lister (2013: 114–115).
71   Cohen (2008: 38).

is explained by speaker hypocrisy. In giving her justification, the kidnapper hypocritically appeals to a value which she is disposed to disregard, namely the value of the child's life.[72]

We can similarly explain why a speaker's moral character makes a difference to public justification on the combined view.[73] Suppose a speaker purports to publicly justify a law to you. As just argued, she therefore purports to address you as a liberal comrade committed to the project of living on terms of mutual respect as free and equal citizens. In virtue of having this commitment, liberal comrades value acting on others only for reasons which they would accept and consider themselves required to impose or advocate for laws only if they are sufficiently confident that they meet this standard. The speaker accordingly purports to be under this requirement. She therefore appeals – if only implicitly – to the value of acting on others only for reasons which they would accept. Yet if the speaker is substantially vicious, she is in fact disposed to disregard this value. Her justificatory act is for this reason hypocritical.

Being addressed in this way disrespects you as free and equal. A successful public justification does not merely aim at communicating reasons why listeners have to accept a law but additionally carries various moral expectations about how listeners ought to act toward the law. They are expected to internalize the law as a constraint on their own behavior, accept liability to coercion and punishment for breaking the law, and hold others accountable to the law when appropriate. However, a substantially vicious speaker is disposed to inappropriately exempt herself from these or similar burdens when it comes to whatever legal or moral requirements that her vice disposes her to disregard.

------

72   Frick (2016: 246).

73   There is a difference between Cohen's kidnapper and a civilly vicious speaker who purports to offer you a public justification. While the kidnapper is directly involved in the situation that his justification is about, this is not necessarily the case for the vicious speaker. Yet, as I argue below, speakers plausibly need not violate a non-involvement condition to lose standing to publicly justify; they merely need to hypocritically appeal to values. This result is supported by the literature on hypocrisy and the standing to blame. An unrepentant thief lacks standing to blame others for stealing, even if she is not involved in their theft. This is because the thief rejects the equality of persons under values relating to property and so forfeits her right to hold people accountable to them (Fritz & Miller 2018: 125). I think that something similar holds for vicious justifiers. Anti-hypocrisy conditions are not wholly uncontroversial – see, for instance, Dover (2019). But even if such conditions turn out not to govern blame or moral justification generally, they still play a central role in a theory of public justification on account of the overriding concern placed on the values of freedom and equality. Vicious speakers purporting to give public justifications disrespect persons as equals because of their hypocrisy and so undermine one of the core values that public justifications are supposed to realize.

By purporting to give a public justification, a vicious speaker treats you as unequal to her under the values to which she appeals, thereby disrespecting you in a way that is incompatible with public justification.

This suffices for a basic defense of a virtue-loaded conception of public justification. Before concluding, I would like to gesture toward some of this conception's implications for public discourse. One might worry that the view that I have sketched threatens the kind of authoritarianism that the Public Justification Principle is supposed to avoid. If civic vice plays the role that I propose, then safeguarding public justification seemingly requires forcing everyone with bad character from political life.

This is not a significant worry because our judgments of which laws to accept are typically based on multiple sources. On my view, if you encounter testimony from two credible speakers, one of whom has substantial civic vice while the other does not, then the law can be publicly justified to you as long as the speaker with adequate character both gives you sufficient, reflectively accessible reason to accept the law and shares a commitment to living on terms of mutual respect. Making this commitment is not difficult; many people have plausibly already done so just in virtue of their upbringing in a pluralistic, liberal democracy. The view I am proposing also is not burdensome for listeners, since it does not forbid at least considering what vicious persons say. Public justification is undermined only in situations where reasonable people must depend on vicious speakers for information without readily available alternatives. My view supports at most arranging testimonial networks to prevent such dependence.

This leads to an alternative worry, namely that my view makes little difference to practice. I should hasten to add that preventing inappropriate testimonial dependence is itself no small difference: it involves reducing the influence of persons with substantial civic vice over public discourse. This serves to illustrate the broader significance of my view. Public justification is an important political good that everyone should do their part to preserve. For listeners, this means that testimony contributing to justifications for laws from persons of substantial civic vice ought to be overlooked in favor of testimony from persons of adequate character. For speakers of substantial civic vice, the result might not typically be a restriction of one's freedom to talk. But it does mean that one's contributions should be canceled out from discourse about laws among the reasonable public. A society's commitment to the ideal of public justification thereby substantially changes how people view themselves as participants in political deliberation.

6   Conclusion

The Dark Knowledge Problem reveals that proponents of public justification have generally overemphasized the role that appealing to reasons plays in discourse contributing to the justification of laws. Although appealing to reasons has its place (namely in securing the epistemically internal conditions for public justification) the civic virtue of speakers makes an essential difference to whether they have standing to publicly justify laws and so forms a constitutive part of the state of public justification itself.

This outcome makes sense. A standard worry about public justification is that it requires us to justify laws to citizens who are morally reprehensible. Burton Dreben is sometimes quoted as responding that proponents of public justification do not have to justify laws to Nazis.[74] My view can be seen as the suggestion that this attitude should be symmetric between the roles of speaker and listener. If proponents of public justification do not have to talk to Nazis, they should not have to listen to them either.

Acknowledgements

Biographical Note

Sean Donahue is a research fellow at the Australian National University.

Works Cited

Applbaum, Arthur Isak (2019). *Legitimacy: The Right to Rule in a Wanton World.* Cambridge, MA: Harvard University Press.

Audi, Robert (1998). "A Liberal Theory of Civic Virtue." *Social Philosophy and Policy* 15 (1): 149–170.

---

74   See Quong (2011: 8) and van Schoelandt (2015: 1032).

Badiola, Cristóbal Bellolio (2018). "Science as Public Reason: A Restatement." *Res Publica* 24: 415–32.

Boettcher, James W. (2012). "The Moral Status of Public Reason." *The Journal of Political Philosophy* 20 (2): 156–177.

Bonjour, Laurence (1980). "Externalist Theories of Empirical Knowledge." *Midwest Studies in Philosophy* 5 (1): 53–73.

Brennan, Jason (2012). "For-Profit Business as Civic Virtue." *Journal of Business Ethics* 106 (3): 313–324.

Budescue, David, Han-Hui Por, Stephen Broomell, and Michael Smithson (2014). "The Interpretation of IPCC Probabilistic Statements around the World." *Nature Climate Change* 4: 508–512.

Cohen, G. A. (2008). *Rescuing Justice and Equality*. Cambridge, MA: Harvard University Press.

Donahue, Sean (2020). "Public Justification and the Veil of Testimony." *The Journal of Political Philosophy* 28 (4): 378–396.

Dover, Daniela (2019). "The Walk and the Talk." *Philosophical Review* 128 (4): 387–422.

Downs, Anthony (1957). *An Economic Theory of Democracy*. New York: Harper & Row.

Ebels-Duggan, Kyla (2010). "The Beginning of Community: Politics in the Face of Disagreement." *Philosophical Quarterly* 60 (238): 50–71.

Feldman, Richard (2003). *Epistemology*. Upper Saddle River: Prentice Hall.

Frick, Johann (2016). "What We Owe to Hypocrites: Contractualism and the Speaker-Relativity of Justification." *Philosophy and Public Affairs* 44 (4): 223–265.

Fritz, Kyle and Daniel Miller (2018). "Hypocrisy and the Standing to Blame." *Pacific Philosophical Quarterly* 99 (1): 118–139.

Gaus, Gerald (2010). "On Two Critics of Justificatory Liberalism: A Response to Wall and Lister." *Politics, Philosophy and Economics* 9 (2): 177–212.

Gaus, Gerald (2011). *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. Cambridge: Cambridge University Press.

Gaus, Gerald and Kevin Vallier (2009). "The Roles of Religious Conviction in a Publicly Justified Polity: The Implications of Convergence, Asymmetry and Political Institutions." *Philosophy and Social Criticism* 35 (1–2): 51–76.

Gettier, Edmund L. (1963). "Is Justified True Belief Knowledge?" *Analysis* 23 (6): 121–123.

Giere, Ronald N. (2009). "Is Computer Simulation Changing the Face of Experimentation?" *Philosophical Studies* 143 (1): 59–62.

Gilbert, Nigel and Klaus G. Troitzsch (2005). *Simulation for the Social Scientist*, 2nd ed. New York, NY: Open University Press.

Goldman, Alvin I. (1999). "Internalism Exposed." *Journal of Philosophy* 96 (6): 271–293.

Gutmann, Amy and Dennis Thompson (2004). *Why Deliberative Democracy?* Princeton: Princeton University Press.

Hartley, Christie and Lori Watson (2018). *Equal Citizenship and Public Reason: A Feminist Political Liberalism*. Oxford: Oxford University Press.

Kappel, Klemens (2021). "Science as Public Reason and the Controversiality Objection." *Res Publica* 27: 619–639.

Larmore, Charles (2008). *The Autonomy of Morality*. Cambridge: Cambridge University Press.

Larmore, Charles (2015). "Political Liberalism: Its Motivations and Goals." In *Oxford Studies in Political Philosophy*: *Vol. 1,* edited by. David Sobel, Peter Vallentyne, and Steven Wall. Oxford: Oxford University Press: 63–89.

Leland, R. J. and van Wietmarschen, Han (2012). "Reasonableness, Intellectual Modesty, and Reciprocity in Political Justification." *Ethics* 122 (4): 721–747.

Leland, R. J. and van Wietmarschen, Han (2017). "Political Liberalism and Political Community." *Journal of Moral Philosophy* 14 (2): 142–167.

Lenz, Gabriel (2012). *Follow the Leader? How Voters Respond to Politicians' Policies and Performance*. Chicago: The University of Chicago Press.

Lister, Andrew (2013). *Public Reason and Political Community*. London: Bloomsbury.

Lupia, Arthur and Matthew McCubbins (1998). *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge: Cambridge University Press.

Mason, Andrew (2007). "Public Justifiability, Deliberation, and Civic Virtue." *Social Theory and Practice* 33 (4): 679–700.

Medvecky, Fabien (2020). "Epistemic Paternalism, Science, and Communication." In *Epistemic Paternalism: Conceptions, Justifications, and Implications,* edited by Guy Axtell and Amiel Bernal (eds.), London: Rowman & Littlefield.

Meyer, Marco, Mark Alfano, and Boudewijn de Bruin (2021). "Epistemic Vice Predicts Acceptance of Covid-19 Misinformation." *Episteme*: 1–22.

Millgram, Elijah (2015). *The Great Endarkenment: Philosophy for an Age of Hyperspecialization*. New York, NY: Oxford University Press.

Nagel, Thomas (1987). "Moral Conflict and Political Legitimacy." *Philosophy & Public Affairs* 16 (3): 215–240.

National Cancer Institute (2007). *Greater Than the Sum: Systems Thinking in Tobacco Control*. Bethesda, MD: US Department of Health and Human Services, National Institutes of Health; Available at: http://dccps.nci.nih.gov/tcrb/monographs/18/index.html.

Oreskes, Naomi (2004). "Science and Public Policy: What's Proof Got to Do With It?" *Environmental Science & Policy* 7: 369–383.

Puerto Rico Governor Resigns After Mass Protests (July 25, 2019). *BBC News*. https://www.bbc.com/news/world-us-canada-49102274.

Quong, Jonathan (2011). *Liberalism without Perfection*. Oxford: Oxford University Press.

Rawls, John (1997). "The Idea of Public Reason Revisited." *University of Chicago Law Review* 64 (3): 765–807.

Rawls, John (1999). *A Theory of Justice: Revised Edition*. Cambridge, MA: Belknap Press.

Rawls, John (2005). *Political Liberalism*. New York: Columbia University Press.

Reid, Andrew (2019). "What Facts Should be Treated as 'Fixed' in Public Justification?" *Social* Epistemology 33 (6): 491–502.

Vallier, Kevin (2019). *Must Politics Be War? Restoring our Trust in the Open Society*. Oxford: Oxford University Press.

Vallier Kevin (2020). "In Defense of Idealization in Public Reason." *Erkenntnis* 85: 1109–1128.

Van Schoelandt, Chad (2015). "Justification, Coercion, and the Place of Public Reason." *Philosophical Studies* 172 (4): 1031–1050.

Waldron, Jeremy (1987). "Theoretical Foundations of Liberalism." *The Philosophical Quarterly* 37 (147): 127–150.