

Why the Convergence Conception of Political Liberalism Cannot Achieve Stability

2020

Abstract

Political liberalism, or public reason liberalism, has taken a decisive turn towards the Convergence Conception of public justification and away from the orthodox Consensus Conception. Convergence theorists argue that public justification should be cashed out as all reasonable people having *some* conclusive reason to endorse coercively enforced issue and context specific moral rules. They argue only the Convergence Conception can achieve a stable political order because, given the nature of deep moral and political disagreement, only it shows reasonable people how to create and maintain a political order. I argue however that, the Convergence Conception cannot achieve what it claims to because it faces the Verbal Agreement Objection. This is because at its core it cashes out the public justification of coercive rules as a series of ‘verbal agreements’ that are inherently unstable. In considering how Convergence theorists might respond, I argue that in avoiding the instability of the Verbal Agreement Objection, theorists are caught in other problems that themselves raise stability issues. I conclude that the Convergence Conception of political liberalism in its current form cannot achieve the stable political order it claims to.

I Introduction

As a theory of political legitimacy, political liberalism, or public reason liberalism, is the view that political principles or rules that are coercively enforced are legitimate if they are ‘publicly justified’ to all reasonable people. When coercively enforced political principles or rules are publicly justified to reasonable people it means they have sufficient moral reason to endorse them. This is the normative standard of ‘public justification’ or “Public Justification Principle” formalised by contemporary political liberals like Gerald Gaus and Kevin Vallier (2009).¹

¹See also Rawls (2005, pp. 69–71, 136, 387–388) Gaus (2016, Ch. 4, 2011, Ch. 5, 1999), Vallier (2019, pp. 4–5) Quong (2011, pp. 180–191).

The central goal of such a theory is to show how reasonable people with profound disagreements about the right and the good can live together in a stable political order in a shared social world.² The ability of a theory to show this is understood along two dimensions: the ability to show how reasonable people can create a political order and the ability to show how they can maintain this political order over time. The ability to show how reasonable people can create a political order amounts to showing how all reasonable people's balance of reasons can provide them sufficient moral reason to coordinate on political principles or rules that are coercively enforced despite their reasonable disagreements. The ability to show how reasonable people can maintain a political order amounts to showing how all reasonable people's balance of reasons can be maintained so they continue to have sufficient moral reason to coordinate on political principles or rules that are coercively enforced despite the endogenous and exogenous forces disturbing their balance of reasons. The forces we are concerned with are limited to, in the case of the endogenous forces, the sphere of activity the political principles or rules themselves permit or encourage. In the case of the exogenous forces we are concerned with the normal changes in circumstances a political society is likely to face. When a theory has the ability to show both of the above things, it shows which political principles or rules allow reasonable people to live in a stable political order. For political liberals, that theory is political liberalism and its core normative standard of 'public justification'.

On the orthodox 'Consensus Conception' of political liberalism public justification is cashed out as the justification of political principles or rules by 'public reasons'.³ These are particular morally conclusive reasons that all reasonable share. The core idea being what explains reasonable disagreements is that reasonable people weigh deliberative considerations (eg. reasons, values etc.) in conflicting ways when making moral and political judgements.⁴ As a result, what justifies coercively enforced political principles or rules is a particular subset of moral reasons – "political values" – that all reasonable people share even if they do weigh them slightly differently.

But some political liberals (and non-liberals) have argued that the Consensus Conception is unstable because the public reasons it identifies are either not actually shared

² As most things in philosophy this view of the structure of political liberalism is somewhat controversial. Some political liberals dissent from this standard reading and argue that in fact political liberalism solely responds to the social fact of disagreement about the good. See Paul Weithman (2015, pp. 75, 83–88, 2010, pp. 319–321) and Jonathan Quong (2011, pp. 36–37, 137–138) on this second interpretation. I do not see things this way side with those hold the more standard reading that political liberalism is an ambitious project responding to fundamental disagreements about justice itself. See Gerald Gaus (2017, pp. 27–30, 2014, pp. 243–248), Kevin Vallier (2019, pp. 5–7), David Thrasher (2018, pp. 399–400), Brian Kogelmann (2017), David Reidy (2007, pp. 250–251), and Burton Dreben (2003, pp. 320–321) for this interpretation.

³ See Rawls (2005), and for those who follow him, Quong (2011), Nussbaum (2011), Lister (2013), Hartley and Watson (2009), Larmore (1999, 1990) and Leland and Wietmarschen (2017).

⁴ See Rawls (2005, pp. 54–58) on the "burdens of judgement", but also Mason (1993, pp. 15, 58, 99–100) and McMahon (2009, pp. 3–4, 55–63, 78, 81) who develop Rawls's explanation into more comprehensive theories.

by all or not actually conclusive and can be overridden by other moral reasons not shared by all reasonable people. As such, they have argued for a turn towards the ‘Convergence Conception’ of public justification.⁵ The core idea being that public justification should be cashed out as the justification of moral rules by *some* morally conclusive reasons. This allows for coercive rules to be justified on the basis of a convergence of potentially conflicting reasons that are each conclusive for a reasonable person within their own point of view. This would vindicate political liberalism as a theory of political legitimacy that can show how reasonable people can achieve a stable political order.

This paper argues the Convergence Conception cannot achieve what it claims to. I argue that it faces what I call the “Verbal Agreement Objection” and that the main responses to this objection also face problems. The basic idea of the objection is that the Convergence Conception is constituted by merely verbal agreements that are highly sensitive to context changes and therefore cannot constitute a political order that can plausibly be maintained by reasonable people. As a result, the Convergence Conception, despite being the best version of the political liberal project, cannot show how reasonable people can achieve a stable political order. To that end, the paper runs as follows. In §2 I explain what the Convergence Conception of political liberalism is and how it differs from the Consensus Conception. In §3 I explain what the Verbal Agreement Objection is. In §4 I consider two ways convergence theorists can respond to the Verbal Agreement Objection by drawing on Ryan Muldoon’s Multi-Perspectival view and Gerald Gaus’s Social Equilibrium view of the mechanism of convergence. The basic idea being that the mechanism of convergence can be spelled out in such away to allow instability in the short-run to ensure stability in the long-run. I then show how both responses fail as well because they open the Convergence Conception to other problems. I conclude that this ultimately shows we have good reason to reject even the most promising conception of political liberalism.

2 The Convergence Conception

The Convergence Conception is best summarised as:

Convergence Conception: A set of issue and context specific moral rules are legitimate if there is a convergence of mutually intelligible reasons amongst all the reasonable people in a society that conclusively justifies endorsing them.

This is the theory of political legitimacy that theorists like Gerald Gaus (2016, 2011)⁶, Kevin Vallier (2019, 2014, 2011; 2009) and Ryan Muldoon (2016) have argued for in

⁵See Thomas Nagel (1987) and Fred D’Agostino (1996) for the earliest explication of the distinction between consensus and convergence in political liberalism.

⁶Strictly speaking Gaus’s theory concerns social morality and the construction of a moral order, with a political order being a tool for constructing and maintaining the moral order (Gaus 2016, pp. 177–187, 206–207, 2011, pp. 460–470, 545–546).

opposition to the Consensus Conception.⁷ Although they all differ on the mechanism of convergence (I will return to this point in §4), the core idea is the same. That is, the public justification of coercively enforced issue and context specific moral rules is a matter of all reasonable people converging in their conclusive justifications through potentially conflicting reasons.

There are two important features that distinguish the Convergence Conception both as a form of political liberalism and as distinct from the orthodox Consensus Conception. Firstly, public justification is cashed out as the convergence of mutually intelligible reasons. That means coercively enforced moral rules are conclusively justified if all reasonable people can find *some* conclusive reason to endorse it and the reason is intelligible as a reason for them by other reasonable people.

The main reason for cashing out public justification in this way is that it avoids a serious objection that plagues the Consensus Conception: the Inconclusiveness Objection.⁸ The core idea is that when reasonable disagreement is taken seriously it will involve ‘deep disagreements’. These are disagreements that are seemingly not about, or at least not only about, the content of the first-order moral or political judgements that people make. Rather they are about people’s “worldviews”, “fundamental commitments” or “perspectives”.⁹ On these disagreements, reasonable people disagree about some of the most fundamental aspects of how they see the world and interpret it, like what counts as evidence or a morally relevant reason. Once these sorts of deep disagreements between reasonable people are taken seriously as genuine disagreements there is no reason to think that there will be a set of reasons that all reasonable people share. As such there is no reason to think there will be coercively enforced principles and rules that can be conclusively justified to all reasonable people on the basis of shared reasons.

But, as political liberals who adopt the Convergence Conception argue, the Inconclusiveness Objection can be avoided if public justification is construed more modestly. That rather than cashing out public justification as justification through a consensus of conclusive reasons, it should be cashed out as justification through *some* mutually intelligible conclusive reasons. This means that each reasonable person can conclusively justify coercive rules through reasons they view as conclusive from their point of view. This means deep disagreements are no impediment to political liberalism.

The second distinguishing feature of the Convergence Conception is that the ob-

⁷ Arguably, Amartya Sen (2010) also advances a sort of Convergence Conception. See for Gaus (2016, pp. 155–163) for discussion on this point.

⁸ See Boettcher (2015, pp. 194–195) for a good discussion of how the inconclusiveness objection, being a variety of the broader incompleteness objection, is similar, but importantly different to the objection put by others in the literature. See also Reidy (2007, p. 261, 2000, pp. 63–70) and Williams (2000). See also Gaus (2011, Ch. 16) on the “Problem of Indeterminacy” and Vallier (2019, pp. 114–115, 2014, p. 215) on the “Anarchy Objection”.

⁹ See Ranalli (2018b, pp. 2–4), Hazlett (2014, pp. 12–13), Kappel (2018, p. 1), and Fogelin (1985) for an overview of deep disagreement in epistemology, and Gaus (2018, p. 647, 2017, 2016, pp. 145–146, 2011, pp. xv–xvii, 2, 38–42) and Muldoon (2016, pp. 1–4) in political philosophy.

jects of justification are not political principles, but context and issue specific rules which all can recognise as *moral rules*. This means that all reasonable people recognise them (for their own reasons) as requirements of justice or morality. Although, these moral rules can then play all sorts of roles from regulating a constitutional framework or the property-rights system, they must be issue and context specific. They are not general principles for how to construct rules or social institutions. Rather they are the moral rules that comprise social institutions themselves. This narrowing of the focus of public justification on issue and context specific moral rules is for two reasons. The first reason is that convergence on general principles is too unlikely given that reasonable people's powers of reason are limited. As Gaus and Vallier say:

Our aim in constructing a justified lawmaking systems is, as far as possible, for acts of legislation to reflect what citizens understand as distinct and manageable issues. Here, as elsewhere, holistic justification is outside the bounds of real human reason. (Gaus 2011, p. 496)

We [convergence theorists] focus on the public justification of moral rules because they're the kind of social practice that can be internalized by most moral agents. Moral life is not based on generic moral principles like Rawls's difference principle, but on local rules governing local behavior. (Vallier 2019, p. 175)

The point here is that people's limited powers of reasoning mean they cannot, given their deep disagreements, adequately decide whether they have sufficient reason to endorse a general political principle against all the other alternatives. Doing this with context and issue specific moral rules is easier and so convergence is more likely.

The second reason for narrowing the focus of public justification is that issue and context specific moral rules are needed for moral rules to do the work of effectively ordering a society so that people who disagree cooperate rather than cheat. As Gaus (2011, p. 113) argues, "group cooperation requires norms or rules that are specific enough in their requirements that cheater detection is highly reliable within the group". As such, rules that do not clearly state what behaviour they require and in the context in which they require them cannot help us detect those that do not cooperate. This would make it difficult for people to actually maintain sufficient reason to endorse moral rules and act according to them.

The point in all of this is that the objects of public justification are moral rules that apply in specific contexts about specific issues like laws or public regulations. They are not general principles encompassing many issues and meant to be applied to various contexts like Rawlsian or Utilitarian principles. Whether general principles of justice can be abstracted out of the set of issue and context specific moral rules is of course another matter. But, on the Convergence Conception whatever these general principles are they are not the objects of public justification.

With those two distinguishing features, political liberals argue that the Convergence Conception avoids the Inconclusiveness Objection and shows how reasonable people can achieve a stable political order.¹⁰ They argue the theory shows how reasonable people can create a political order despite their deep disagreements, since all that's required to publicly justify issue and context specific moral rules is reasonable people to have *some* mutually intelligible conclusive reason to endorse it even if they all disagree about what that reason is (Vallier 2019, pp. 89–98; Muldoon 2016, pp. 62–65; Gaus 2011, pp. 391–408). The idea is that when the normalisation on reasons is lowered to the level of 'mutual intelligibility' far more possible reasons can conclusively justify a coercively enforced moral rule including the ones people disagree about in deep disagreements. This shows how reasonable people can have sufficient moral reason to endorse issue and context specific moral rules despite their deep disagreements.

Political liberals then argue that this political order can be maintained because the convergence is not over any specific set of reasons. This means that when reasonable people's set of moral reasons changes or their weighing of them changes, this will simply change the set of issue and context specific moral rules that are converged upon (Vallier 2019, pp. 116–121, 2014, pp. 164–170; Gaus 2011, pp. 310–321). This means there is no point at which the publicly justified set of rules are oppressive or justify anarchic rebellion since whenever they become so, they will no longer be publicly justified since at least one reasonable person will not have conclusive reason to endorse them. Since the Convergence Conception holds on to the core idea that only publicly justified moral rules can be coercively coordinated upon, the political order is always maintained.

3 Verbal Agreement Objection

Whilst the Convergence Conception does well at avoiding the Inconclusiveness Objection, some political liberals in favour of the Consensus Conception have pointed out that the way it avoids the objection is problematic.¹¹ They argue that because of shifts in people's beliefs and the changing circumstances for a society, the nature of convergent agreements and the objects of those agreements means that the political order reasonable people create following the Convergence Conception is coincidental. Even though it is not entirely unlikely given people's particular beliefs (ie. they could not have converged on anything else), it is lucky that they converged on a particular moral rule because they could easily have converged on a different one with slightly different contexts or beliefs. This initial luck of convergence means it lacks a kind of normative stability. This is because, on the Convergence Conception, a political order is created by reason-

¹⁰Interestingly, Boettcher (2015, pp. 201–204) demurs and argues some form of inconclusiveness remains a problem because of the controversial notion of coercion assumed by convergence theorists. I leave aside Boettcher's objection and side with Gaus's (2014) response to similar objections as adequate.

¹¹See Hartley and Watson (2018, pp. 59–61).

able people agreeing to the issue and context specific moral rules rather than general political principles. This means that when people's beliefs change or when the context changes the agreement will break down.

But, why should convergence theorists be particularly worried by this? After all, it is not new that people change their beliefs or that circumstances change. That is the nature of social and political life. Whilst this is true, I think there is something more to the objection that should worry convergence theorists, namely that convergent agreements are *disposed* to break down rather than coincidental in their creation.

The basic idea is that an upshot of explaining deep disagreements as genuine disagreements – the sort of disagreement that motivated the Convergence Conception to begin with – is that at least some of the agreements that constitute a convergence on issue and context specific moral rules are 'mere verbal agreements'. These are agreements grounded purely in the difference between reasonable people's views on the relevant deliberative considerations for forming some belief and the context they happen to deliberate in. However, this means these agreements are highly sensitive to context change. This is because the agreement is grounded in the convergence of distinct deliberative considerations and the specific contexts in which those considerations justify endorsing the moral rule. Slight context changes will mean that at least one person's considerations will not recognise the issue and context specific moral rule in question as a moral rule to endorse. The result of this is that at least some of the agreements that constitute a political order on the Convergence Conception are *disposed* to break down. Aspects of the political order that require endurance or quick institution building are by their very nature likely to become illegitimate. The political order is, in short, unstable. All reasonable people do not have sufficient moral reason to support it over time and therefore the theory cannot show them how to achieve a stable political order. I call this the Verbal Agreement Objection.

That is all fairly abstract. To get a better sense of how the Verbal Agreement Objection let us start by understanding how explaining deep disagreements as genuine has the upshot I've claimed above. As I have mentioned, reasonable disagreements that are deep disagreements, seem to be about something other than the first-order moral and political judgements people make. They seem to be about some matter at a deep level of thought involving how people see the world and how to interpret it. This is what, as I mentioned earlier, theorists mean by saying deep disagreements seem to be about "worldviews", "perspectives" or "fundamental commitments". For instance, consider the following case:

Nationalisation: Bryan and Elizabeth are discussing the economic structure their society ought to have. Elizabeth judges their society ought to nationalise, at the very least, some key industries because a society's productive capacity being for the mutual benefit of all is what justice requires. Bryan judges their society ought not to nationalise any industries because

protecting people's natural moral right to their body and private property is what justice requires. This is because he believes natural moral rights are what matter, benefiting everyone with society's productive capacity is irrelevant and has nothing to do with what justice requires. On the contrary, Elizabeth believes materially benefiting people is what matters, natural rights to private property are irrelevant and have nothing to do with what justice requires.

In *Nationalisation* Bryan and Elizabeth do not disagree about the economic structure justice requires in virtue of their beliefs about what justice requires. Rather they disagree about what matters for even thinking and making judgements about what justice requires.

However, this raises the question of what makes deep disagreements like *Nationalisation* genuine disagreements? Typically, if disagreements do not seem to be about a conflict in the content of the beliefs or judgements that concern the first-order matter at hand, then they are not genuine disagreements. They are verbal disagreements. This is because, a genuine disagreement is one that involves a conflict in mental content (eg. beliefs, desires, plans, etc.) such that one person holds content P, the other holds Q, and P and Q cannot both be true.¹² It seems in a case like *Nationalisation*, whatever Bryan and Elizabeth are disagree about it is not about the content of their beliefs about the first-order matter at hand: the economic structure justice requires. Yet, intuitively, cases of deep disagreement like *Nationalisation* do seem to be genuine disagreements. It does seem like Bryan and Elizabeth do conflict over some mental content that matters or is related in some way to their conflicting judgements about the economic system justice requires. It seems that they would carry on disagreeing even after they explained themselves to each other, and that the resolution of their disagreement would practically matter for how they led their lives.

Fortunately theorists have proposed various explanations of how deep disagreements are genuine disagreements. Some theorists explain it epistemically by citing conflicts over special types of beliefs they hold (ie. "hinge beliefs", "entitlement claims", "substantive epistemic principles", "framework propositions").¹³ Other theorists explain it semantically or conceptually by citing conflicts over the content of one's conceptual or evaluative framework.¹⁴ Regardless of which explanation one adopts, they all share an underlying structure. They all work by arguing that deep disagreements actually involve two conflicts. The first, the conflict over the first-order matter is a proxy for or indirect representation of, a second deeper conflict. The second deeper conflict, to

¹²See Frances (2014, Ch. 1) and Gibbard (2003, Ch. 4) for this basic way of thinking about genuine disagreement.

¹³See Ranalli (2018a), Hazlett (2014), Pritchard (2018), Kappel (2018), Adams (1985) and Fogelin (1985).

¹⁴See Gaus (2018, pp. 648–650, 2016, pp. 43–45, 51–56, 2011, pp. 276–279) and Muldoon (2016, pp. 80–85) who build on the idea of "diverse perspectives" developed by Scott E. Page (2007, pp. 30–31) in complexity theory.

summarise the various explanations, is over what reasonable people believe, categorise or interpret as the morally relevant deliberative considerations for making moral and political judgements. This means they can disagree, not because they hold conflicting first-order moral and political beliefs, but because they hold conflicting beliefs, categorisations or interpretations of the morally relevant deliberative considerations for forming those beliefs. For example, *Nationalisation* is explained as Bryan and Elizabeth's conflict over the economic system that justice requires being a proxy for their conflict in views of the morally relevant deliberative considerations for making judgements about what justice requires. It is the second conflict that makes a deep disagreement like *Nationalisation* a genuine disagreement.

So far so good. With an explanation of why deep disagreements are genuine, the Consensus Conception faces the Inconclusiveness Objection and motivates the Convergence Conception. However, an upshot of explaining deep disagreements as genuine is that it also explains how reasonable people who hold conflicting views of the morally relevant deliberative considerations for making moral and political judgements, can also genuinely *agree* if their views diverge conveniently enough to support practically *compatible* moral and political judgements in the context in which they deliberate. Once deep disagreements are genuine such agreements are likewise genuine. This would be a case of a 'mere verbal agreement'.¹⁵

But notice mere verbal agreements are nothing but an example of the idea of convergence where reasonable people endorse an issue and context specific moral rule for potentially diverse conflicting reasons. Mere verbal agreements are convergences grounded in people's conflicting views about the morally relevant deliberative considerations for making moral and political judgements. This means that on the Convergence Conception public justification is, at least some of the time, a series of agreements where reasonable people are lucky enough that their views of the morally relevant deliberative considerations for making moral and political judgements diverge conveniently enough that they lead them to make moral and political judgements that are practically *compatible* in the context in which they live.

For example, Alice and Beth could agree to an issue and context specific moral rule that says resources ought to be distributed according to talents in their society. Alice believes the rule is what justice requires because it helps individuals and society flourish and, individuals and society flourishing is the only relevant consideration for a rule about distributing resources. On the other hand, Beth believes the rule is what justice requires because she believes it gives people what they morally deserve and moral desert is the only relevant moral consideration for a rule about a distributing resources. In such a case Alice and Beth's agreement would be a merely verbal agreement because they agree on the rule wholly in virtue of the differences in their views of the morally

¹⁵See Chalmers (2011, pp. 525–526) for this way of construing verbal agreements. See also Ballantyne (2016) on construing verbal agreements through the same answer satisfying two different questions.

relevant deliberative considerations for making moral and political judgements. It is only in virtue of the distinct and unique considerations they view as morally relevant for deliberating about moral rules that enables them to agree.

The problem with a merely verbal agreement like Alice and Beth's is that when it is over a issue and context specific moral rule it is highly sensitive to changes in context. It is highly sensitive to changes in the facts that constitute the moral rule (ie. changes in what a rule permits or prohibits as a matter of fact). For instance, a slight change in what counts as a talent, suppose Beth discovers people have a natural talent to cheat others, might cause the agreement to break down because cheating others does not promote moral desert. As such Beth would not see the rule as conclusively justified anymore given what she views as morally relevant in her social world, namely moral desert and nothing else.

These sorts of context changes can come in roughly two varieties: exogenous and endogenous changes. The first involves changes in the facts that constitute a moral rule that are caused by facts external to the rule. Rapid technological advancements and scientific, or perhaps even philosophical discoveries, can change what states of affairs a moral rule applies to or what states of affairs fall under a deliberative consideration. This can cause agreements struck earlier contexts to breakdown under these new and unforeseen context.¹⁶ This is worrying because what complex and diverse societies require are an enduring social institution that can decide on new claim disputes and policy problems.

The second variety of context change involves a change in the facts that constitute a moral rule caused by the implementation of the rule itself. The enforcement of some rule on a complex system like a society's economic, which is also interconnected with other complex systems, will change the system itself.¹⁷ This will then change what states of affairs a person counts as evidence for future judgements about that very rule and therefore change the deliberations of the people that it must be justified to. In this new context, the earlier agreement on the rule will break down purely because of its own enforcement and therefore the rule will no longer be publicly justified. Once again, this seems worrying because it would mean that there is no publicly justified rule about some issue when some rule is sorely required.

¹⁶If one doubts context changes like this can have such consequences, consider the rapid escalation of the coronavirus pandemic of 2019-2020 and the way many countries without enduring social institutions were not able to form new agreements to combat the health and economic crisis. See <https://nymag.com/intelligencer/2020/03/coronavirus-paid-leave-health-care-trump.html>

¹⁷See Gaus (2016, pp. 230–231) for someone who seems to acknowledge this form of instability.

4 Mechanism of Convergence Response

One way convergence theorists are likely to respond to the Verbal Agreement Objection is by arguing the instability identified by the Verbal Agreement Objection is a red herring. They will likely argue that such instability is actually a feature of the Convergence Conception. A convergent agreement breaking down in the sort of ways I have argued is good because it signals that the moral rules that were publicly justified are no longer justified and ought not be obeyed or coercively enforced. To illustrate this point, convergence theorists will likely draw on the distinction between ‘stability’ as ordinarily understood and ‘robustness’. Stability as ordinarily understood is the tendency of a system to return to the same unique equilibrium point, and robustness is the tendency to return to some equilibrium point (Vallier 2019, pp. 193–195; Gaus 2016, pp. 230–237). As such, the Convergence Conception shows reasonable people can maintain a political order in the robustness sense. If we model the mechanism of convergence in the right way we will see how a political order can adapt to the short-term changes in context so that it continually returns to a convergent agreement. In what follows I lay out two prominent ways this response are likely to be made – The Multi-Perspectival Bargaining View and The Social Equilibrium View – and argue that both have problems.

4.1 The Multi-Perspectival Bargaining View

Ryan Muldoon has recently argued that the mechanism of convergence should be seen as the outcome of what he calls “Multi-Perspectival Bargaining”.¹⁸ On this view, a convergence on a set of issue and context specific moral rules is the result of reasonable people bargaining about the public rules (which specify rights), that ought to govern their shared social world. It is “multi-perspectival” because it takes place between reasonable people who hold conflicting political beliefs *and* conflicting views of the morally relevant deliberative considerations for forming those beliefs (Muldoon 2016, pp. 24, 64–64).

Muldoon (2016, pp. 77–84) argues that despite holding different and conflicting perspectives, reasonable people can bargain over what rules ought to govern their shared social world because the underlying physical states of affairs are the same. Although each person has a different perspective of these states of affairs, each recognises the stakes of having their opponent’s rules governing over them according to their perspective. This allows for a bargain that results in a “joint individual justification” on rules over the shared social world.

Muldoon (2016, p. 83) argues the bargaining will take place much like bargaining in a marketplace where two parties agree to a mutually advantageous price for an exchange.

¹⁸Note Muldoon’s account is novel and detailed. Explaining it in full with all its arguments is beyond the scope of this paper. Therefore in what follows I focus purely on the aspects that would serve as a response to the Verbal Agreement Objection.

In the case of issue and context specific moral rules reasonable people weigh up how much a set of rules conforms to *their* moral and political beliefs against how much it violates it. This weighing up of the costs and benefits of the rules determines the “price” parties are willing to pay to agree to them. Muldoon (2016, pp. 72–77) argues that people can bargain over rules in this way because the object of the bargaining are rules that specify rights according to a “social conception” of rights. According to this social conception, rights come in different varieties (eg. positive and negative rights, rivalrous and non-rivalrous, and excludable and non-excludable) and as “bundles of allowances and guarantees” that can be broken up into “many affiliated allowances for action or guarantees that may be considered at least somewhat independently of each other”. As such, two people who do not share a perspective on how to categorise what is morally relevant can bargain their way to a set of rights over their shared social world. They are able to agree on the states of affairs they are discussing (even if these states of affairs are a small overlap of what each person sees) and on the nature of the rules that will apply to them, but not why the other person thinks the rules are right or correct.

Muldoon (2016, pp. 110–111) concludes that just like marketplaces, the bargaining will contain a measure of instability as people constantly reassess their rights.¹⁹ But, this instability merely reflects people deciding whether they are made better off by a new bargain or staying in the old one. Muldoon’s account of the mechanism of convergence is compelling and does provide a plausible picture of why the Verbal Agreement Objection might not be a serious problem. If a society has, as Muldoon argues, reason to value the adaptation to changes in context and perspectives to strike new bargains, then the instability of mere verbal agreements does not seem all that worrying. But, there are three serious problems with the Multi-Perspectival Bargaining view of the mechanism of convergence that undermine its ability to show a political order can be created let alone maintained.

The first is the Problem of Mutual Advantage. As Gaus (2016, pp. 170–171) argues, The Multi-Perspectival Bargaining way of understanding convergence makes achieving a stable political order hinge on what is mutually advantageous to the bargainers when they make an agreement. This means there is nothing in the Multi-Perspectival Bargaining view that prevents someone from backing out of an agreement on rules as soon as a rule or the “price” they paid for it is no longer advantageous to them. In fact the view encourages it as a way for people to strike new bargains. But, as Gaus (2016, pp. 170–171) rightly points out one of the core ideas of *reasonable* people trying to agree to rules is that they seek rules that bind people despite how it might advantage them in the future. This is part of the idea of endorsing a rule as a rule that justice requires. In short, the Multi-Perspectival view seems to have an odd idea of the type of rules reasonable people

¹⁹See Muldoon (2016, pp. 90–91, 102–107) on the three principles that a bargain will have to satisfy for his specific argument for how reasonable people will reassess old bargains and make news because they value diversity and social experimentation.

are trying to propose and agree to.

The second problem is the Problem of Calculation. As Gaus (2016, pp. 172–173) The Multi-Perspectival view underestimates the complexity of predicting changes in context. This complexity undermines the way reasonable people are supposed to conclusively justify the set of coercive rules by negotiating their way to an agreement. It seems almost impossible for someone to calculate the costs and benefits of striking a bargain when they cannot predict how striking a particular bargain may change the context which will itself require a new bargain. As an example Gaus (2016, p. 172) refers to rules concerning immigration which affect the entry of new reasonable people into a society and therefore the make up of the set of perspectives that the constituency of reasonable people hold. But, even in a ‘closed society’, rules concerning education, or healthcare directly affect the context which will change the set of perspectives and therefore require a new bargain. Bargainers would need to predict all of this to have some idea of judging what rules they can agree to live under.

But aside from those two problems have already been canvassed in the literature there is, I submit, a much more worrying problem that has been overlooked. This is what I call the Problem of Uni-Perspectival Rights. Recall, that one of the things that is supposed to make the Multi-Perspectival Bargaining view work is that the coercive rules that people bargain over specify rights according to a “social conception” of rights.

The problem with this is precisely that it is a particular conception of rights and not a multi-perspectival one. There is no reason to think reasonable people will share such a conception because, as we saw in §3, reasonable can have genuine deep disagreements. This means that when they are endorsing a rule, there is no reason to think they will see it as specifying the kind of right that Muldoon assumes.²⁰ Their views of the morally relevant deliberative considerations for making moral and political judgements need not employ Muldoon’s conception of rights.

For instance, reasonable people may not view a right as merely an “allowance” and “guarantee”. They may employ a more or less sophisticated idea of what a right even is. Even if there is little scope for divergence here, there will certainly be a lot of disagreement about the precise features of rights. For instance, reasonable people who believe all rights are political and institutionally defined or believe some (albeit a large portion) of rights are grounded in people’s interests, will, or given by God will not agree with Muldoon’s social conception of rights let alone all its features. For example, it is not clear what Muldoon’s view has to say when one bargainer suggests breaking up the ‘freedom of conscience’ such that it does not apply to confessions to priests, but another bargainer rejects that particular freedom *can* be broken up in such a way. The point here is that the sort of bargaining that allows Muldoon’s view to work relies on a conception of rights that will be the subject of the cases of deep disagreement that make

²⁰This problem also plagues others in the public reason liberalism tradition. See Chung (2019) on liberal rights and Chung and Kogelmann (2020, p. 849) on jurisdictional rights for examples.

up reasonable disagreement about justice itself.

All in all, the Multi-Perspectival Bargaining view is an inadequate response to the Verbal Agreement Objection. It is replete with problems running from pure practical problems of what the view requires and encourages, to a deeper problem to relying on a unique perspective on rights. This undermines the mechanism's ability to show a political order can be created let alone maintained.

4.2 The Social Equilibrium View

The most well developed alternative to the Multi-Perspectival Bargaining view of the mechanism of convergence is the Social Equilibrium view developed by Gerald Gaus and also used by Kevin Vallier.²¹ On this view, a convergence on issue and context specific moral rules is an “equilibrium of social norms” (Vallier 2019, pp. 192–195; Gaus 2016, 2011, pp. 434–443). The basic idea is that social norms – albeit a special set of social norms – constitute a “social-morality” that already exists in societies with reasonable disagreement, and that convergence should be modelled as an equilibrium of these social norms.

The Social Equilibrium view's model of convergence relies on two key ideas. The first is that, drawing on Cristina Bicchieri's work on social norms, we ought to see the coercive rules that reasonable people converge on as social norms.²² The idea is that social norms are social rules in that they set a general standard of behaviour that is supposed to be followed by society and there is a practice of criticism and punishment for those who violate them. In a simple sense, social norms track people's reactive attitudes to certain actions in a society. Social norms with distinct moral content are social rules of a certain kind: moral rules. The distinct moral content is an ongoing mutual recognition of a structure of reciprocal deontic obligations and expectations. This structure involves, roughly, six features which make moral rules the sort of rules that can be publicly justified. The rules are sufficiently general, intelligible to all, validate claims and resolve conflicts, place requirements on behaviour not mere guidelines, are endorsable by others in different social positions, and proposed as being for the good of all such that all can reasonably internalise them (Gaus 2011, pp. 294–303). Moral rules are then as Gaus (2011, pp. 181–182) says “justified deontic requirements that a large part of the population intends (at least conditionally) to follow and are actually conformed to by a large part of a group's members”. That is what makes the social norms that are moral rules capable of being the object of public justification.

When a scheme of moral rules are publicly justified, which means all reasonable

²¹Much like Muldoon's, note that Gaus's account is novel and rich in detail. As such, explaining it in full with all its arguments is beyond the scope of this paper. Therefore in what follows I focus purely on the aspects that would serve as a response to the Verbal Agreement Objection.

²²See Bicchieri (2016, 2006) on the general empirical account, and Gaus (2016, pp. 211–215, 2011, pp. 163–182) and Vallier (2019, pp. 30–36) on its implementation in political liberalism.

people have sufficient moral reason to endorse them, they constitute a social-morality which establishes the “moral order” or “public moral constitution” of a society. Given this moral order is publicly justified, a political order of coercive laws and institutions is legitimate if it enforces the moral order (Vallier 2019, Ch. 6, 7; Gaus 2016, pp. 206–207, 2011, pp. 449–470). This is because the political order is merely an enforcement of the moral rights that the moral rules establish.

The second key idea in the Social Equilibrium view concerns how reasonable people can identify the moral rules that all have sufficient moral reason to endorse. The idea is to see the mechanism of convergence on a set of moral rules as an equilibrium that emerges out of the actual path-dependent social interactions of reasonable people (Vallier 2019, pp. 33–36, 110–113; Gaus 2016, pp. 223–226, 2011, pp. 321–322, 389–408). Convergence – and hence public justification – is then an emergent phenomenon that arises out of reasonable people’s everyday social interactions where social norms are agreed to and reassessed continuously.

The process of convergence is then modelled in two stages. In the first stage, reasonable people, with their conflicting political beliefs, and conflicting views of the morally relevant deliberative considerations for forming those beliefs, rank social rules which they judge as strictly better than no authoritative rule at all. The set of social rules that all reasonable people rank in this way constitutes the *socially eligible set* because no rule in that set is judged as worse than no authoritative rule at all on a particular issue or as Gaus (2011, p. 322) says, “rules that all have reasons to endorse as authoritative”.²³ A reasonable person endorses a rule as authoritative if two conditions are met. The first condition is that they recognise that those proposing a rule have some intelligible reason for endorsing it *as a moral rule* given the definition above (Gaus 2016, pp. 212–215, 2011, p. 322). This means the socially eligible set will not include mere social conventions that people might have some reason to follow, but do not count as moral rules (eg. rules of etiquette). The second condition is that they have *some* reason to internalise it as a moral rule with its normative requirements and expectations. As Gaus (2011, p. 333) says, “The set of rules that all Members of the Public have some reason to accept as authoritative yields what I have called a socially eligible set.” The important point here of course is not that reasonable people have sufficient moral reason to endorse any particular rule compared to the other rules in the eligible set. Rather it is that, as Gaus (2011, pp. 323, 325) says, the socially eligible set is a modest conclusion such that if a rule is not in the socially eligible set it “fails to be publicly justified in a strong sense” such that reasonable judge those rules to “palpably fail to adequately perform their tasks.” This will leave a set of moral rules about some specific issue that all reasonable people have some reason to endorse as authoritative. This means that they are willing to hold others accountable

²³I leave aside the additional step of arriving at the *optimal* socially eligible set which includes only the rules that are not pareto dominated by any other rules because convergence theorists admit for most cases the socially eligible set is enough and the additional step is irrelevant for the objection I press (Gaus 2016, p. 215).

for complying with even though each reasonable person ranks the rules within that set in different ways. This is the socially eligible set.

After forming a socially eligible set, the second stage of public justification involves reasonable people converging on a unique moral rule – which means they find a rule they all have conclusive reason to endorse – in the socially eligible set by interacting with each other based on those rules (Gaus 2016, pp. 223–226, 2011, Ch. 19).²⁴ As such, convergence is achieved by reasonable people acting on moral rules according to 1) what moral rules *best* satisfy their view of the morally relevant deliberative considerations for forming political beliefs and 2) the extent to which acting on the moral rules other reasonable people are coordinating on allows them to reap the benefits of having social interactions where they can “respect their equality and moral freedom” (Gaus 2011, pp. 398–399). The benefits are that they will be able to make morally authoritative demands on others that are publicly justified. Reasonable people will then converge because there are two forces that are pushing people into agreement. One force is the satisfaction of their view of the morally relevant deliberative considerations and another force is the benefit of endorsing rules that others also endorse so they can enjoy social interactions that respect other’s equality and moral freedom. The point is reasonable people will be able to weigh these two considerations in such a way that they do not merely hold out for rules that *most* satisfy what they view as the morally relevant deliberative considerations, but reconcile in favour of rules that satisfy these considerations as far as is possible whilst also allowing them to have social interactions with publicly justified rules. The crucial normative upshot of this whole process is, as Gaus (2011, p. 414) argues, all reasonable people can come to have sufficient moral reason to endorse moral rules even though they started by disagreeing on what they took to be the ideal moral rules. In short, the model shows how a set of coercive rules can be in equilibrium between all reasonable people who reconcile according to their view of the morally relevant deliberative considerations and the benefit of having social interactions with publicly justified rules.

The result of all of this is that convergence on a set of issue and context specific moral rules need not make use of the idea of bargaining or some particular conception of rights. Therefore it can avoid all three of the problems faced by the Multi-Perspectival Bargaining view. Rather, as Gaus (2011, pp. 402–403, 2011, pp. 225–226) argues, it can be an emergent outcome of people’s path-dependent social interactions and revisions of social norms. As such, it can be an emergent equilibrium of social norms that can be legitimately explicated, enforced and maintained by coercive political power because

²⁴Strictly speaking, Gaus (2016, pp. 198–202, 2011, Ch. 4) and Vallier (2019, pp. 202–214) argue that reasonable people must look to narrow the socially eligible set before this step. This involves honing on the moral rules that need to be justified first by focusing on rights that need to be settled to preserve people’s status and capabilities as moral agents, and “jurisdictional rights” that establish private spheres of conduct where each reasonable person’s perspective is morally authoritative. I skip over this step because they are orthogonal to how the Social Equilibrium view is a response to the Verbal Agreement Objection.

they are the moral rules, in virtue of being in equilibrium, that are publicly justified. This is how reasonable people can achieve a stable political order given reasonable disagreement. On this view, reasonable people create a political order by freely endorsing moral rules and then constructing coercive institutions to coercively enforce them. This occurs because of the way moral rules are taken to be informal social norms that people coordinate on through social interactions they learn or adopt. An equilibrium of such moral rules will constitute an equilibrium point that describes the publicly justified moral constitution of a society. The political order is then constructed to enforce this moral constitution to either keep the unreasonable from straying or to ensure that all understand the costs of violating it.

This political order is then maintained by a feedback loop between the underlying moral order of social norms that fall in and out of convergence and the coercive laws in place. This solves the Verbal Agreement Objection because it shows how changes in context will be met with changes in people's reciprocal normative obligations and expectations which will shift the equilibrium of social norms. This is because the underlying moral order, rather than being a set of centralist principles that govern all human interactions, are a network of rules that concern different behaviours at different levels of generality and in different parts of a society. This feature of social norm networks allow for changes in context to be met with localised violations where people attempt to coordinate on a new rule in the socially eligible set. It allows for a practice of moral reform and criticism so that reasonable people can move to a new equilibrium of moral rules in the same process that the political order was created (Gaus 2016, pp. 226–230, 2011, pp. 433–443). This means reasonable people's interpersonal exchanges and moral deliberations pushed them back into an equilibrium. Given only moral rules that are in equilibrium can be coercively enforced, the political order will follow the changes in the moral order. With new coercive laws and institutions being constructed to adapt to changes in social norms. On this view, the Verbal Agreement Objection is a red herring because mere verbal agreements are simply shifts in the equilibrium of social norms that constitutes a publicly justified moral constitution which is then enforced by a political order of coercive laws and institutions. The socially eligible set provides a set of rules that reasonable people can constantly reflect on and deliberate about endorsing. As a result, any changes in context are part and parcel of how reasonable people converge on coercive rules.

4.3 The Integrity Objection Returns

I concede that the Social Equilibrium view does avoid the Verbal Agreement Objection. It offers a convincing account of how the mechanism of convergence can be understood to avoid the sort of instability I argued was entailed by the Convergence Conception. But, I submit, all this comes at a cost. To show how a political order can be maintained over time, the view relies on the coercive social pressure of people coordinating on social

norms to publicly justify rules. Therefore, the view faces a version of the Integrity Objection. It requires reasonable people to coordinate on moral rules to avoid coercive social pressure rather than to conform to the morally relevant deliberative considerations they view as integral to their view of justice and their social world.

The Integrity Objection is a well canvassed objection against the Consensus Conception of political liberalism.²⁵ The idea being that requiring non-liberals to only justify political principles and decisions according to the reasons they share with other reasonable people requires them to split their persona between acting as political liberals in the political domain and acting as they truly think is just in the private domain. This infidelity to their true character, plans and beliefs means political liberalism attacks their integrity as reasonable people. The Convergence Conception is, rightly, seen to avoid this sort of objection.²⁶ After all, the Convergence Conception on the Social Equilibrium view allows reasonable people to use the reasons they see as valid reasons. It allows them to act according to their comprehensive conceptions of justice all the time.

But, I propose, there is a version of the Integrity Objection that operates at the level of a person's views of the morally relevant deliberative considerations which the Social Equilibrium view cannot avoid.²⁷ To understand how this version of the Integrity Objection works recall again how the Social Equilibrium view proposes reasonable people come to endorse a particular moral rule within the socially eligible set. According to the Social Equilibrium view, reasonable people endorse a single rule within the socially eligible set by evaluating the rules it contains according to what they view as the morally relevant deliberative considerations for making moral and political judgements *and* the benefits of coordinating on rules that others are coordinating on. This means that as more people endorse a rule the benefits of also endorsing that rule increase and therefore the reasons to endorse the rule becomes weightier and eventually conclusive.

The problem with this is that for some reasonable people the overriding reason that will conclusively justify a moral rule is the coercive social pressure of other reasonable people endorsing and acting on a social norm. They face others holding them accountable in their day to day social interactions. This is because at least for some reasonable people, the moral rule that is endorsed by others is the one that they rank as the one that least satisfies the deliberative considerations they view as morally relevant for making moral and political judgements about justice. Although they can see the rule as better than no rule at all, it is *barely* better. This, I submit, threatens their integrity. It threatens the beliefs, categorisations and interpretations of the morally relevant deliberative considerations which are integral to them. This is because achieving a stable political order, depends on the overriding weight of others endorsing a rule as reason to also endorse it. It does not depend on reasons related to a rule conforming accurately to

²⁵See Vallier (2012, pp. 156–160), Eberle (2002, pp. 143–151), and Wolterstorff (1997, p. 105) for a good overview.

²⁶See Eberle (2011, pp. 291–293) and Vallier (2012, pp. 161–164) on this point.

²⁷See Waldron (2015) for the closest version of the objection I press.

what is morally relevant in their social world let alone what justice requires. Achieving a stable political order depends on at least some reasonable people foregoing acting on rules that are more integral to them.

This is all very abstract. To get a sense of the version of the Integrity Objection I am pushing here, take a concrete example like healthcare. Consider an issue and context specific moral rule \mathcal{M} : Individuals and groups in the provision of healthcare have a right to discriminate against individuals on the basis of congenital illnesses and for no other reason. Suppose \mathcal{M} is in the socially eligible set, meaning that all reasonable people have some reason to endorse it as authoritative according to the deliberative considerations they view as morally relevant for making judgements about what justice requires. Now, suppose Gordon, as part of the minority in this society, evaluates \mathcal{M} as better than no rule at all because he thinks no authoritative rule at all on this issue would simply lead to even worse discrimination for individuals looking to be cared for. This would affect people finding healthcare for even common non-congenital illnesses. But, Gordon evaluates \mathcal{M} as the worst possible morally authoritative rule in the socially eligible set (ie. it is closer to being ruled out by the deliberative consideration he views as morally relevant than any other rule). On the deliberative considerations that Gordon views as morally relevant, he has much weightier reason to endorse all the other moral rules in the socially eligible set. Perhaps these are rules that each permit discrimination to a small degree, but not on a whole category of illnesses. Suppose now that a large majority in Gordon's society, contrary to Gordon, endorses \mathcal{M} . What is Gordon to do? What moral rule within the socially eligible set does Gordon have sufficient moral reason to endorse and coordinate on?

On the Social Equilibrium view, Gordon ought to endorse and act on \mathcal{M} purely on the fact that the majority in his society endorse \mathcal{M} . If it weren't for that fact, he would have sufficient moral reason to endorse some other rule within the eligible set. This is not to restate the Social Equilibrium view's mechanism of convergence. Rather it is to point out that the reasonable people who are left in the minority as their society slowly converges on a single moral rule in the socially eligible set, face a stark choice. Either endorse a rule purely on the basis that others are endorsing it *and* that the rule is in their eligible set, or endorse on the basis of what they have most moral reason to endorse and face the social consequences of dissenting.

Given the way all this is supposed to be occurring through informal social interactions, it shows how creating a moral order depends on people endorsing a rule to avoid the coercive social pressure of the majority in their society acting on a moral rule they judge as barely better than no rule at all. The sole reason why Gordon will converge is that a sufficiently large number of people endorse a social norm and that joining them is more beneficial than not. Gordon ought to endorse a moral rule not because his beliefs, categorisations and interpretations of the morally relevant deliberative considerations for making moral and political judgements are accurate or true, but because he faces "resentment," "indignation" and the "guilt" of acting on a moral rule that the

majority are not converging on (Gaus 2016, p. 181, 2011, pp. 409–412). This threatens his integrity. He cannot achieve a stable political order by finding moral rules that he has sufficient moral reason to endorse according to his view of the morally deliberative considerations. Rather he has to also consider the fact that the majority of reasonable people in his society endorse some other moral rule.

One thought might be, so what? What is so bad about the threat to integrity in this way? The main problem is that it generates its own source of instability. Reasonable people are likely to mistrust and hate their fellow reasonable citizens if their integrity is constantly under threat by the social pressure to endorse a social norm. This is no philosopher's flight of fancy. There is considerable evidence from political psychology that the more disagreement people face in their everyday interactions, the more likely they are to become ambivalent about or disengage from political procedures and the social practice of discussing political issues.²⁸ The worry in all of this is that there is good reason to think that reasonable people will turn their back on trying to find the moral rule that all can endorse within the socially eligible set. There is good reason to think they will not accept the coercive power of social pressure since it is necessarily beyond their control, but will accept a potentially oppressive arrangement since they can at least hope to control it if I can persuade enough people. Any equilibrium change is then disposed to fail on the Social Equilibrium view because reasonable people are justified in holding out for a moral rule that is more supported by what they view as the morally relevant deliberative considerations. This is because they will judge the process by which social norms become publicly justified moral rules less controllable or lacking in compromise from all parties than some other method. This is a severe cost for the Social Equilibrium view. It would undermine the core mechanism by which a political order is maintained. Constant shifts in the equilibrium will appear oppressive because the path to justifying them involves succumbing to the costs of transgressing them as social norms.

This puts convergences theorists in a bind. On the one hand they face the Verbal Agreement Objection if they choose to be neutral on the mechanism of convergence. On the other they can accept that although the Social Equilibrium view is the best version of the Convergence Conception it comes at the cost of threatening integrity. On either side they face the problem of not being able to maintain a political order.

One response convergence theorists might make is to say that the version of the Integrity Objection I have made fundamentally misunderstands what it means for a moral rule to be in the socially eligible set. They might argue it does not mean the moral rule is supported by at least some of the considerations a reasonable person views as morally relevant. Rather it means that the moral rule is *conclusively* justified by some deliberative consideration for every reasonable person. This means it does not make

²⁸See Diana Mutz (2006), Huckfeldt (2008; 2004), Wojcieszak (2012; 2011), Nir (2011) and McClurg (2006), and Barnidge (2017) for further evidence of this in social media networks.

sense to say that people have more or less weighty reason to endorse the rules relative to each other in the socially eligible set.²⁹ On this view, a rule being in the socially eligible set does not mean that reasonable people merely have *some* reason to endorse it as I suggested earlier. Rather it means they have a conclusive reason – which establishes its moral authority – to endorse it.

This may well be a plausible way to read the Social Equilibrium view. However, it would open the Convergence Conception up to the return of the Inconclusiveness Objection. Given the lesson of deep disagreements is that people will have conflicting views of the morally relevant deliberative considerations, it is to be expected that reasonable people will interpret and evaluate certain rules and rights in deeply irreconcilable ways. A stronger threshold for when a person grants authority to a moral rule will yield an empty socially eligible set. There is no reason to think reasonable people will strike upon a set of rules that all have conclusive moral reason to endorse.

Another response convergence theorist might make is that, given deep disagreements are genuine, any theory of political legitimacy that employs convergence is going to involve some threat to integrity. The depth of reasonable disagreement means that any way that a theory of political legitimacy shows how a reasonable people can conclusively justify political principles or rules is going to justify principles or rules that are some reasonable person's least justifiable option. There is no way around this. So, this is not a cost unique to the Social Equilibrium view.

I think this is far too premature. The real lesson is that the Social Equilibrium view is the best version of the Convergence Conception. But, it comes at a cost. It threatens the integrity of reasonable people and so risks undermining the core mechanism of creating and maintaining a political order in their eyes. This motivates retaining the core idea of convergence, but exploring other types of theories. After all, political liberalism is not the only way to theorise about how reasonable people can have sufficient moral reason to coordinate on political principles or rules that are coercively enforced. For instance, contemporary political realism has recently been argued for as a theory of political legitimacy that can account for the true depth of moral diversity in modern societies. Political realists argue that rather than public justification, what reasonable people need is a normative standard that is context-dependent, and justifies an attitude of practical acceptance of principles or rules rather than endorsement which requires internalising them. Evaluating this as a potential alternative is beyond the scope of this paper, but it suffices to illustrate where convergence theorists might explore to avoid the instability of public reason liberalism.

²⁹See Gaus (2011, p. 425) for this possible reading.

5 Conclusion

In this paper I have argued that the Convergence Conception of political legitimacy cannot achieve a stable political order because it cannot show how reasonable people can maintain the order it creates. I argued it faces the Verbal Agreement Objection because an upshot of explaining how deep disagreements are genuine disagreements was that some convergent agreements would be “mere verbal agreements”. In short, I argued that convergent agreements on issue and context specific moral rules between reasonable people who conflict in their views of what the morally relevant deliberative considerations are for making moral and political judgements will be highly sensitive to changes in context. As such, at least some of the political order the Convergence Conception purports to create cannot be maintained in the face of even slight changes in social contexts.

I then argued that although there are two ways convergence theorists can model the mechanism of convergence to respond to the Verbal Agreement Objection, both of them face their own objections. I argued the Multi-Perspectival View is not an adequate version of the response because it encourages people to back out of agreements when it no longer favours them, requires complex calculations of the effect of a bargain, and requires a particular conception of individual rights that will itself be the subject of reasonable disagreement. As such, it cannot show how reasonable people can create a political order.

I then argued that the Social Equilibrium View is also not an adequate version of the response because it faces a version of the Integrity Objection. It relies on a mechanism that threatens the views of the morally relevant deliberative considerations for making moral and political judgements that are integral to reasonable people. Therefore it generates its own sources of instability for maintaining a political order.

I concluded that although the Social Equilibrium view is the best version of the Convergence Conception it comes at the cost of threatening reasonable people’s integrity. This undermines the core mechanism by which it shows how reasonable people can maintain a political order. As such, the Convergence Conception is caught between the instability of the Verbal Agreement Objection, and the instability of the Integrity Objection. Therefore, the charge that the Convergence Conception cannot show reasonable people how to create a stable political order stands.

References

- Adams, David. “Knowing when Disagreements are Deep”. In: *Informal Logic* 25.1 (1985), pp. 65–77.
- Ballantyne, Nathan. “Verbal Disagreements and Philosophical Scepticism”. In: *Australasian Journal of Philosophy* 94.4 (2016), pp. 752–765.

- Barnidge, Matthew. "Exposure to Political Disagreement in Social Media Versus Face-to-Face and Anonymous Online Settings". In: *Political Communication* 34.2 (2017), pp. 302–321.
- Bicchieri, Cristina. *Norms in the Wild: How to Diagnose, Measure and Change Social Norms*. Cambridge University Press, 2016.
- *The Grammar of Society: The Nature and Dynamics of Norms*. Cambridge University Press, 2006.
- Bicchieri, Cristina and Peter McNally. "Shrieking Sirens: Schemata, Scripts, and Social Norms. How Change Occurs". In: *Social Philosophy and Policy* 35.1 (2018), pp. 23–53.
- Boettcher, James W. "Against the Asymmetric Convergence Model of Public Justification". In: *Ethical Theory and Moral Practice* 18.1 (2015), pp. 191–208.
- Chalmers, David. "Verbal Disputes". In: *Philosophical Review* 11.4 (2011), pp. 515–566.
- Chung, Hun. "The Impossibility of Liberal Rights in a Diverse World". In: *Economics and Philosophy* 35.1 (2019), pp. 1–27.
- Chung, Hun and Brian Kogelmann. "Diversity and rights: a social choice-theoretic analysis of the possibility of public reason". In: *Synthese* 197 (2020), pp. 839–865.
- D'Agostino, Fred. *Free Public Reason: Making It Up As We Go*. Oxford University Press, 1996.
- Dreben, Burton. "On Rawls and Political Liberalism". In: *The Cambridge Companion to Rawls*. Ed. by Samuel Freeman. Cambridge University Press, 2003.
- Eberle, Christopher. "Consensus, Convergence, and Religiously Justified Coercion". In: *Public Affairs Quarterly* 25.4 (2011), pp. 281–303.
- *Religious Conviction in Liberal Politics*. Cambridge University Press, 2002.
- Fogelin, Robert. "The logic of deep disagreements". In: *Informal Logic* 7.1 (1985), pp. 1–8.
- Frances, Bryan. *Disagreement*. Polity Press, 2014.
- Gaus, Gerald. "Is Public Reason a Normalization Project? Deep Diversity and the Open Society". In: *Social Philosophy Today* 33.1 (2017), pp. 27–52.
- "Reasonable Pluralism and the Domain of the Political: How the Weaknesses of John Rawls's Political Liberalism Can be Overcome by a Justificatory Liberalism". In: *Inquiry* 42.2 (1999), pp. 259–284.
- "The Complexity of a Diverse Moral Order". In: *The Georgetown Journal of Law & Public Policy* 16.1 (2018), pp. 645–680.
- *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. Cambridge University Press, 2011.
- "The Turn to a Political Liberalism". In: *A Companion to Rawls*. Ed. by Jon Mandle and David A. Reidy. Wiley Blackwell, 2014.
- *The Tyranny of the Ideal: Justice in a Diverse society*. Princeton university Press, 2016.

- Gaus, Gerald and Kevin Vallier. "The roles of religious conviction in a publicly justified polity: The implications of convergence, asymmetry and political institutions". In: *Philosophy Social Criticism* 35.1–2 (2009), pp. 51–76.
- Gibbard, Allan. *Thinking How to Live*. Harvard University Press, 2003.
- Hartley, Christie and Lori Watson. *Equal Citizenship and Public Reason: A Feminist Political Liberalism*. Oxford University Press, 2018.
- "Feminism, Religion, And Shared Reasons: A Defense Of Exclusive Public Reason". In: *Law and Philosophy* 28.5 (2009), pp. 493–536.
- Hazlett, Allan. "Entitlement and Mutually Recognized Reasonable Disagreement". In: *Episteme* 11.1 (2014), pp. 1–25.
- Huckfeldt, Robert, Paul E. Johnson, and John Sprague. *Political Disagreement: The Survival Of Diverse Opinions Within Communication Networks*. Cambridge University Press, 2004.
- Huckfeldt, Robert and Jeanette Morehouse Mendez. "Moths, Flames, and Political Engagement: Managing Disagreement within Communication Networks". In: *The Journal of Politics*, 70.1 (2008), pp. 83–96.
- Kappel, Klemens. "Higher Order Evidence and Deep Disagreement". In: *Topoi* (2018), pp. 1–12. URL: <https://doi.org/10.1007/s11245-018-9587-8>.
- Kogelmann, Brian. "Justice, Diversity, and the Well-Ordered Society". In: *The Philosophical Quarterly* 67.269 (2017), pp. 663–684.
- Larmore, Charles. "Political Liberalism". In: *Political Theory* 18.3 (1990), pp. 339–360.
- "The Moral Basis Of Political Liberalism". In: *The Journal of Philosophy* 96.12 (1999), pp. 599–625.
- Leland, R. J. and Han van Wietmarschen. "Political Liberalism and Political Community". In: *Journal of Moral Philosophy* 14.2 (2017), pp. 142–167.
- Lister, Andrew. *Public Reason and Political Community*. Bloomsbury, 2013.
- Mason, Andrew. *Explaining Political Disagreement*. Cambridge University Press, 1993.
- McClurg, Scott D. "Political Disagreement in Context: The Conditional Effect of Neighborhood Context, Disagreement and Political Talk on Electoral Participation". In: *Political Behaviour* 28.4 (2006), pp. 349–366.
- McMahon, Christopher. *Reasonable Disagreement: A Theory of Political Morality*. Cambridge University Press, 2009.
- Muldoon, Ryan. *Social Contract Theory for a Diverse World: Beyond Tolerance*. Routledge, 2016.
- Mutz, Diana C. *Hearing the Other Side: Deliberative versus Participatory Democracy*. Cambridge University Press, 2006.
- Nagel, Thomas. "Moral Conflict and Political Legitimacy". In: *Philosophy & Public Affairs* 16.3 (1987), pp. 215–240.
- Nir, Lilach. "Disagreement and Opposition in Social Networks: Does Disagreement Discourage Turnout?" In: *Political Studies* 674–692.3 (2011), pp. 149–160.

- Nussbaum, Martha C. "Perfectionist Liberalism and Political Liberalism". In: *Philosophy & Public Affairs* 39.1 (2011), pp. 3–45.
- Page, Scott E. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press, 2007.
- Pritchard, Duncan. "Wittgensteinian Hinge Epistemology and Deep Disagreement". In: *Topoi* (2018), pp. 1–9. URL: <https://doi.org/10.1007/s11245-018-9612-y>.
- Quong, Jonathan. *Liberalism Without Perfection*. Oxford University Press, 2011.
- Ranalli, Chris. "Deep disagreement and hinge epistemology". In: *Synthese* (2018), pp. 1–33. URL: <https://doi.org/10.1007/s11229-018-01956-2>.
- "What is Deep Disagreement?" In: *Topoi* (2018), pp. 1–16. URL: <https://doi.org/10.1007/s11245-018-9600-2>.
- Rawls, John. *Political Liberalism: Expanded Edition*. New York: Columbia University Press, 2005.
- Reidy, David A. "Rawls's Wide View Of Public Reason: Not Wide Enough". In: *Res Publica* 6.1 (2000), pp. 49–72.
- "Reciprocity And Reasonable Disagreement: From Liberal To Democratic Legitimacy". In: *Philosophical Studies* 132.2 (2007), pp. 243–291.
- Sen, Amartya. *The Idea of Justice*. Penguin Books, 2010.
- Thrasher, John and Kevin Vallier. "Political Stability in the Open Society". In: *American Journal of Political Science* 62.2 (2018), pp. 398–409.
- Vallier, Kevin. "Convergence and Consensus in Public Reason". In: *Public Affairs Quarterly* 25.4 (2011), pp. 261–280.
- *Liberal Politics and Public Faith: Beyond Separation*. Routledge, 2014.
- "Liberalism, Religion And Integrity". In: *Australasian Journal of Philosophy* 90.1 (2012), pp. 149–165.
- *Must Politics Be War?: Restoring Our Trust in the Open Society*. Oxford University Press, 2019.
- Waldron, Jeremy. "Isolating Public Reasons". In: *Rawls's Political Liberalism*. Ed. by Thom Brooks and Martha C. Nussbaum. Columbia University Press, 2015.
- Weithman, Paul. "Legitimacy and the Project of Political Liberalism". In: *Rawls's Political Liberalism*. Ed. by Thom Brooks and Martha C. Nussbaum. Columbia University Press, 2015.
- *Why Political Liberalism?: On John Rawls's Political Turn*. Oxford University Press, 2010.
- Williams, Andrew. "The Alleged Incompleteness of Public Reason". In: *Res Publica* 6.2 (2000), pp. 199–211.
- Wojcieszak, Magdalena E. "Pulling Toward or Pulling Away: Deliberation, Disagreement, and Opinion Extremity in Political Participation". In: *Social Science Quarterly* 92.1 (2011), pp. 206–225.

- Wojcieszak, Magdalena E. and Vincent Price. "Perceived Versus Actual Disagreement: Which Influences Deliberative Experiences?" In: *Journal of Communication* 62.3 (2012), pp. 418–436.
- Wolterstorff, Nicholas. "The Role of Religion in Decision and Discussion of Political Issues". In: *Religion in the Public Square: The Place of Religious Convictions in Political Debate*. Ed. by James P. Sterba and Rosemarie Tong. Rowman & Littlefield, 1997.