

12-12-2018

The Fragility of Convergence: Public Reason, Political Liberalism and Stability

Paul Fryfogle

Follow this and additional works at: https://scholarworks.gsu.edu/philosophy_theses

Recommended Citation

Fryfogle, Paul, "The Fragility of Convergence: Public Reason, Political Liberalism and Stability." Thesis, Georgia State University, 2018.
https://scholarworks.gsu.edu/philosophy_theses/247

This Thesis is brought to you for free and open access by the Department of Philosophy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Philosophy Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

THE FRAGILITY OF CONVERGENCE: PUBLIC REASON, POLITICAL LIBERALISM
AND STABILITY

by

PAUL FRYFOGLE

Under the Direction of Christie Hartley, PhD

ABSTRACT

John Rawls's shift from *A Theory of Justice* to *Political Liberalism* was prompted by his dissatisfaction with *Theory*'s account of stability. Rawls's later account of stability places the idea of public reason at its center. On one influential reading, Rawls models stability as a mutual assurance game, wherein stability is provided by assurance of commitment to the shared conception of justice. In recent years, a growing number of convergence theorists have argued that Rawls's assurance mechanism, in the form of public reason, fails in societies characterized by pervasive disagreement. These theorists propose convergence as offering superior assurance. This thesis argues that the stability convergence offers is fragile due to the very pluralism with which convergence theorists are rightly concerned. Moreover, convergence does not do well at explaining why citizens, understood as politically free and equal, wish to cooperate on fair terms.

INDEX WORDS: public reason, political liberalism, consensus, convergence, stability

THE FRAGILITY OF CONVERGENCE: PUBLIC REASON, POLITICAL LIBERALISM
AND STABILITY

by

PAUL FRYFOGLE

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2018

Copyright by
Paul Fryfogle
2018

THE FRAGILITY OF CONVERGENCE: PUBLIC REASON, POLITICAL LIBERALISM
AND STABILITY

by

PAUL FRYFOGLE

Committee Chair: Christie Hartley

Committee: William Edmundson

Andrew Jason Cohen

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

September 2018

DEDICATION

To my mother

ACKNOWLEDGEMENTS

I want first to thank Christie Hartley, who supervised my thesis and from whom I have learned a great deal, for providing extensive and insightful comments on countless drafts of this project, and for seeing it through each stage with patience, generosity, and dedication. My thanks also go to Andrew Jason Cohen and William Edmundson for their generosity in providing many thoughtful criticisms and suggestions. Their comments were all invaluable and I have tried to accommodate as many of them as possible. I bear the responsibility for any remaining shortcomings. I wish also to thank Linh Mac and Luke Pizzato for many friendly discussions on topics in this paper and on political philosophy in general. Both have been a great source of challenge and insight. Many thanks to Sandy Dwyer for her helpful comments on an early version of this paper, and for her mentorship with all things graduate school and teaching while at Georgia State University. Special thanks are due also to long-time academic mentor Marcus Arvan for going far beyond the call of any professional duty in offering invaluable guidance throughout my academic career. I am grateful also to Marcus, Liv Coleman, and Steven Geisz for their profound influences on me as an undergraduate. I am grateful to Tyler Vanwulven and Jessica Scott for their ongoing encouragement, discussion, and friendship. Finally, I am grateful to Chelsea Margueritte Walker for her unmatched love and support.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	V
1 INTRODUCTION.....	1
2 STABILITY AND THE IDEA OF PUBLIC REASON	3
2.1 The Problem of Stability	4
3 CONSENSUS VS. CONVERGENCE.....	9
3.1 Justificatory Reasons: Shareability and Intelligibility	10
4 CONVERGENCE DISCOURSE AND THE PROBLEM OF MUTUAL ASSURANCE	14
5 THE FRAGILITY OF CONVERGENCE	16
6 CONCLUDING REMARKS	25

1 INTRODUCTION

John Rawls's shift from *A Theory of Justice* to *Political Liberalism* was prompted by his dissatisfaction with *Theory*'s account of stability.¹ Rawls's later account of stability places the idea of public reason at its center. On one influential reading, Rawls models stability as a mutual assurance game, wherein stability is provided by assurance of commitment to the shared conception of justice.² Contract theorists have long been concerned to show how a stable political order is possible despite deep and pervasive disagreement. The story is a familiar one from Hobbes to Rawls: stability requires that each of society's members has the assurance that they will not be taken advantage of should they choose to cooperate, that is, members want to do right by their fellow citizens only if they believe that their fellow citizens will do right by them. For Rawls, the threat to stability arises on the assumption that each member *wants* to act justly but needs the assurance that others will do the same, even when temptations arise from within

¹ Some scholars, most notably Brian Barry, Samuel Freeman, and Paul Weithman, highlight Rawls's dissatisfaction with, in particular, the argument in chapter IX of *Theory* as prompting his move to political liberalism. See Brian Barry, *John Rawls and the Search for Stability*, ETHICS 890 (1995); SAMUEL FREEMAN, *Congruence and the Good of Justice*, in THE CAMBRIDGE COMPANION TO RAWLS 278 (Samuel Freeman ed., 2003); PAUL WEITHMAN, WHY POLITICAL LIBERALISM?: ON JOHN RAWLS'S POLITICAL TURN 68-96 (2010) [hereinafter WEITHMAN, WHY POLITICAL LIBERALISM]. Others, most notably Burton Dreben and Allen Buchanan, highlight the shift in Rawls's concern from providing an account of justice to that of legitimacy, while Weithman, on the other hand, has insisted that what originally prompted Rawls's move to political liberalism was his interest in showing that justice as fairness—including *both* of Rawls's principles—would be stable for the right reasons. See BURTON DREBEN, *On Rawls and Political Liberalism*, in THE CAMBRIDGE COMPANION TO RAWLS 316 (Samuel Freeman ed., 2003) [hereinafter DREBEN, *On Rawls*]; PAUL WEITHMAN, RAWLS, POLITICAL LIBERALISM AND REASONABLE FAITH 91 (2016) [hereinafter WEITHMAN, REASONABLE FAITH]. Further still Dreben, more controversially, has thought that what prompted Rawls's move was seeing the first two-thirds of *Theory* as contradicting the last third. See DREBEN, *On Rawls*, at 316. *But cf.* WEITHMAN, REASONABLE FAITH 85-111 (2016). The interpretation of Rawls's dissatisfaction with part III of *Theory* that I am assuming is sufficiently general to align with the main thrust of most of these accounts, and in places where it diverges I make note.

² WEITHMAN, WHY POLITICAL LIBERALISM, *supra* note 1, at 326-335; Weithman has since been joined by Stephen Macedo and Gillian Hadfield.

the “self-interested point of view.”³ Rawls famously remarks, “even with a sense of justice men’s compliance with a cooperative venture is predicated on the belief that others will do their part.”⁴ The question is whether from the self-interested point of view—when the gains of injustice outweigh the desire to act justly—the rational course of action will be to free-ride on others’ cooperative efforts. Without the assurance that others will maintain their sense of justice rather than free-ride, or in any other way ignore their sense of justice, defection becomes the dominant strategy. This is the basic problem of mutual assurance.

In recent years, “convergence” liberalism has established itself as the main theoretical competitor to Rawlsian “consensus” liberalism. Some convergence theorists argue that Rawls’s solution to the assurance problem fails because Rawls’s assurance mechanism, in the form of public reason, is fragile in societies characterized by reasonable pluralism. In short, these theorists argue that when Alf and Betty engage in public reason, Alf cannot infer from Betty’s adherence to its guidelines that Betty will actually choose to maintain her sense of justice rather than defect.⁵ In other words, Betty’s public display of constrained deliberation does not provide Alf assurance that she will actually maintain her sense of justice. These theorists propose convergence as offering a superior assurance mechanism. I argue, however, that the stability convergence offers is fragile due to the very pluralism with which convergence theorists are rightly concerned. Moreover, convergence does not do well at explaining why citizens, understood as politically free and equal, wish to cooperate on fair terms. This paper is organized as follows. Section I discusses the relationship between public reason and stability. Section II

³ JOHN RAWLS, A THEORY OF JUSTICE 295 (rev. ed. 1999) [hereinafter RAWLS, THEORY].

⁴ *Id.* at 336.

⁵ Gaus, *Tale of Two Sets*, *supra* note 40, at 317-18; Kogelmann and Stich, *When Public Reason Fails Us*, *supra* note 42, at 5. Thrasher and Vallier, *The Fragility of Consensus*, *supra* note 41, at 943-946.

sketches the basic features of convergence *à la* its chief architects Gerald Gaus and Kevin Vallier, on the one hand, and Rawls's more familiar consensus account of public reason liberalism, on the other. Section III canvasses the assurance problem and convergence theorists' attempts to solve it. Finally, Section IV argues that convergence's assurance mechanism renders stability fragile and unrealistic in deeply diverse societies.

2 STABILITY AND THE IDEA OF PUBLIC REASON

The idea of public reason is primarily a view about the justification of principles and policies in a liberal society. In particular, it is the reason of equal citizens in a liberal constitutional democracy when they act as a collective to exercise political power. For Rawls, public reason is a set of guidelines stipulating that when making arguments and decisions about constitutional essentials and matters of basic justice, citizens and public officials are to rely on political values drawn from a political conception of justice. But public reason also has other important functions. On one influential reading of Rawls's later thought—most notably the interpretations of Paul Weithman,⁶ Stephen Macedo,⁷ and Gillian Hadfield⁸—the assurance problem is solved by public reason. Understood in this way, we can view public reason as playing both a normative *and* a practical role. First, there is the “moral, not [] legal, duty”—the duty of civility—to be able to explain how principles and policies that members advocate are justified by the political values of public reason.⁹ And second, there is the practical role of securing mutual assurance by citizens signaling their commitment to the political conception over their private

⁶ WEITHMAN, WHY POLITICAL LIBERALISM, *supra* note 1, at 326-335.

⁷ Stephen Macedo, *Why Public Reason? Citizens' Reasons and the Constitution of the Public Sphere*, in SSRN ELECTRONIC JOURNAL 32-36 (2010) [hereinafter Macedo, *Why Public Reason*].

⁸ Gillian Hadfield and Stephen Macedo, *Rational Reasonableness: Toward a Positive Theory of Public Reason*, LAW AND ETHICS OF HUMAN RIGHTS 8-15, 46 (2012).

⁹ JOHN RAWLS, POLITICAL LIBERALISM 216 (expanded ed., 2005) [hereinafter RAWLS, POLITICAL LIBERALISM].

interests. When citizens adhere to the guidelines of public reason while deciding on constitutional essentials and matters of basic justice, they signal their allegiance to the political conception over their comprehensive doctrines, and when they do not adhere to the guidelines of public reason, they signal their allegiance to their comprehensive doctrines over the political conception.¹⁰ Exactly how citizens do this is the subject of much of this paper.

2.1 The Problem of Stability

Rawls tells us in the 1993 preface to *Political Liberalism* that “to understand the nature and extent of the differences [between *Political Liberalism* and *Theory*], one must see them as arising from trying to resolve a serious problem internal to justice as fairness, namely from the fact that the account of stability in part III of *Theory* is not consistent with the view as a whole.”¹¹ The kind of stability Rawls is referring to is “inherent” stability, which he contrasts with the imposed stability of the Hobbesian sovereign.¹² Rawls remarks that when inherent stability obtains, “inevitable deviations from justice are effectively corrected or held within tolerable bounds by

¹⁰ I do not mean to suggest, nor does Weithman et al., that public reason should be understood as a type of practical agreement. Rather, public reason understood as playing two roles—a normative and a practical role—is what Weithman’s reading suggests (and Macedo and Hadfield are explicit about its “practical” role), and Vallier, John Thrasher, Brian Kogelmann, and Stephen G. W. Stich (whose views I canvass below) follow Weithman in viewing public reason as playing both roles.

¹¹ *Id.* at xv-xvi.

¹² See JOHN RAWLS, LECTURES ON THE HISTORY OF POLITICAL PHILOSOPHY 78-79 (Samuel Freeman ed., 2008); JOHN RAWLS, *The Sense of Justice*, in COLLECTED PAPERS 103-104 (Samuel Freeman ed., 1999); RAWLS, THEORY, *supra* note 3, at 125, 436; See also Samuel Freeman, *Political Liberalism and the Possibility of a Just Democratic Constitution*, 69 CHI. KENT L. REV. 623-627 (1994). See also WEITHMAN, WHY POLITICAL LIBERALISM, *supra* note 1, at 43-51. *But cf.* George Klosko, *Rawls, Weithman, and the Stability of Liberal Democracy*, RES PUBLICA 241-244 (2015). For more on the distinction between inherent and imposed stability, see Edward McClennan, *Justice and the Problem of Stability*, PHILOSOPHY AND PUBLIC AFFAIRS 7-8 (1989). The Rawls of *Political Liberalism* spoke of showing “stability for the right reasons” rather than “inherent stability”. Nevertheless, Rawls’s underlying concern was the same for both. See also WEITHMAN, WHY POLITICAL LIBERALISM, *supra* note 1, at 67. Thus I shall use “inherent stability” and “stability for the right reasons” interchangeably throughout this paper.

forces within the system.”¹³ When Rawls says that he is trying to show the inherent stability of justice as fairness, he means that he is trying to show that the basic institutions informed by it, taken together, generate their own support.

Recall that part III of *Theory* treats the problem of stability in two parts.¹⁴ The first part, discussed in chapter 8 of *Theory*, argues that members of the well-ordered society would normally acquire an effective sense of justice. Rawls sketches a three-stage process of moral development effected by just institutions to show how this is so.¹⁵ Just institutions—crudely put—shape the character of those who live under them, such that they answer in kind to the just actions of others. In the second part, Rawls argues that each member of the well-ordered society would, from the viewpoint of “full deliberative rationality”¹⁶—taking account of all her desires—judge that it would be good for her to maintain her sense of justice. Because the character formation needed for stability would be effected by institutions informed by justice as fairness, Rawls thought that justice as fairness—once institutionally realized and publicized—would stabilize itself.

However, Rawls’s argument for inherent stability, and in particular for the second of these two parts, depends upon a solution the assurance problem. But a well-ordered society is one whose basic structure is regulated, and is known to be regulated, by principles of justice and is inhabited by members with an effective sense of justice—it is a society in which “everyone accepts and knows that the others accept the same principles of justice,”¹⁷ where “acceptance”

¹³ RAWLS, *THEORY*, *supra* note 3, at 401.

¹⁴ *Id.* at 397.

¹⁵ *Id.* at 405-434.

¹⁶ *Id.* at 417 (“It is the plan that would be decided upon as the outcome of careful reflection in which the agent reviewed, in the light of all the relevant facts, what it would be like to carry out these plans and thereby ascertained the course of action that would best realize his most fundamental desires.”).

¹⁷ *Id.* at 4.

implies a disposition, which each knows that everyone has, to act on those principles. We might think that, ipso facto, the stability problem has been overcome since each member has a sense of justice and a well-ordered society has full compliance. But this is not so. Rawls tells us that problems remain.¹⁸ Even with a sense of justice members need to be reasonably sure that if they are to act justly others will do the same. Rawls says that citizens may come to “lack full confidence in one another” and “may suspect that some are not doing their part, and so they may be tempted not to do theirs.”¹⁹ This is the heart of the mutual assurance problem for stability.

To the extent that what follows appears to rely too heavily on Weithman’s reading of Rawls’s move to political liberalism—specifically, to Weithman’s emphasis on the role that public reason plays in solving the assurance problem—I offer the following two pleas. First, nearly all the current scholarship on the relationship between public reason and mutual assurance situates itself in relation to Weithman’s canonical reading. And second, it is Weithman’s reading that the convergence theorists I discuss in this paper are working from, and as such it does not seem wholly inappropriate to work from the same.

A distinctive feature of Weithman’s interpretation of Rawls’s move to political liberalism is that he sees the problem of stability from part III of *Theory* as, in effect, the same as in *Political Liberalism*. Namely, the problem is one of showing how acting justly is the best reply to others acting justly. This is Rawls’s Nash Equilibrium model, according to which stability is a “condition of general equilibrium” in which “everyone knows that everyone else acts justly, and each replies to the justice of others by being just himself.”²⁰

Rawls offers different solutions to the assurance problem throughout his career, and these

¹⁸ *Id.* at 295-296, 435.

¹⁹ *Id.* at 211.

²⁰ WEITHMAN, WHY POLITICAL LIBERALISM, *supra* note 1, at 44.

differences plausibly explain the changes from *Theory* to *Political Liberalism*. The Rawls of *Theory* attempted to solve the assurance problem by showing that a congruence relationship obtains between the concepts of justice and goodness, that is, that having an effective sense of justice is a good for each member of a well-ordered society. Rawls argued that from the standpoint of “full deliberative rationality,” members want to live up to ideals of personal conduct, friendship, and association included in justice as fairness as part of their good.²¹ Rawls thought that the value members in the well-ordered society attached to the realization of these ideals would “normally outweigh” temptations to defect.²² Whatever members think they might gain from free-riding, by, say, evading their taxes, or acting in any other way contrary to their sense of justice, is normally outweighed by the good of realizing these ideals. Seeing their endorsement of their sense of justice as central to their good, each member is motivated to comply with the public conception and knows that everyone else is similarly motivated. Hence, common knowledge of compliance obtains and mutual assurance is preserved, or, as Weithman puts it, “Each member judges...that her balance of reasons tilts in favor of *maintaining* her desire to act from the principles of justice as a highest-order regulative desire in her rational plan, when the plans of others are similarly regulated.”²³

Dissatisfied with this account, the Rawls of *Political Liberalism* introduces the idea of an

²¹ Specifically, Rawls says that the desire to be just is “practically identical” with the desire to attain four ends. RAWLS, *THEORY*, *supra* note 3, at 499. These are the desire to (i) “express [their] nature as free moral persons”; (ii) avoid the psychological costs of hypocrisy and deception; (iii) “protect[] in a natural and simple way the institutions and persons [they] care for”; (iv) participate in forms of social life that “bring to fruition [their] latent powers.” *Id.* at 499-501.

²² WEITHMAN, *WHY POLITICAL LIBERALISM*, *supra* note 1, at 85. Here Weithman draws our attention to the long footnote at the end of JOHN RAWLS, *Political Not Metaphysical*, in *COLLECTED PAPERS* 414 n.33 (Samuel Freeman ed., 1999). And later, in *Political Liberalism*, Rawls tells us that the values and ideals of justice as fairness “*normally outweigh* whatever values oppose them.” RAWLS, *POLITICAL LIBERALISM*, *supra* note 8, at 155 (emphasis added).

²³ WEITHMAN, *WHY POLITICAL LIBERALISM*, *supra* note 1, at 63.

overlapping consensus, wherein each affirms the political conception for reasons internal to their own reasonable comprehensive doctrines. In *Political Liberalism* as in *Theory*, Rawls tells us that “stability involves two questions.”²⁴ The first is whether members of the well-ordered society would acquire a sense of justice. Rawls continued to think they would. The second is whether members would judge that they should maintain their sense of justice.²⁵ On the new account as on the original one, stability depends upon members wanting to live up to certain ideals that are only realized when they maintain their sense of justice. But instead of stability depending upon their wanting to live up to the set of *ethical* ideals assumed in *Theory*, the Rawls of *Political Liberalism* argues that stability depends upon their wanting to live up to *political* ideals. In effect, then, overlapping consensus establishes what *Theory*’s congruence argument failed to.²⁶

But just as in *Theory* so in *Political Liberalism*, the decision to maintain one’s sense of justice depends upon a solution to the assurance problem. The Rawls of *Political Liberalism* implies that the public knowledge of the existence of an overlapping consensus would solve it,

²⁴ RAWLS, POLITICAL LIBERALISM, *supra* note 8, at 141.

²⁵ On this account, the political conception would stabilize itself in the second way because “a reasonable and effective political conception may *bend comprehensive doctrines toward itself*.” *Id.* at 246 (emphasis added). The political conception, when institutionalized, encourages an overlapping consensus of reasonable comprehensive doctrines.

²⁶ As Weithman painstakingly argues, Rawls introduced the idea of an overlapping consensus to establish what Weithman calls *Theory*’s “*Nash Claim*” (C_N):

Each member of the WOS [well-ordered society] judges, *from within the thin theory of the good*, that her balance of reasons tilts in favor of maintaining her desire to act from the principles of justice as a highest-order regulative desire in her rational plans, when the plans of others are similarly regulated. WEITHMAN, WHY POLITICAL LIBERALISM, *supra* note 1, at 63.

Weithman tells us that Rawls’s political liberal solution to *Theory*’s problem then moves us to the *Nash Claim* of *Political Liberalism* (C_N^*):

Each member of the WOS [well-ordered society] judges, *from within her comprehensive view*, that the balance of reasons tilts in favor of maintaining her desire to live up to the values and ideals of justice as fairness, at least when others live up to those values as well. *Id.* at 275.

thereby securing stability.²⁷ In other words, citizens must not only affirm the political conception on the basis of their own comprehensive doctrines, they must know that others do so too. But as is the focus of much of Weithman’s study, “[W]hen the concepts used by the very fully comprehensive doctrines are so different from those of justice as fairness [or any one of a class of liberal conceptions], how can it be known to take part in an overlapping consensus? How is public knowledge of an overlapping consensus possible?”²⁸ Weithman argues that the prominence of public reason in *Political Liberalism* is explained in part by the need to solve this very problem.²⁹

3 CONSENSUS VS. CONVERGENCE

As I emphasized in the previous section, Rawls’s account of public reason is advanced as part of an argument for the stability of a well-ordered society. But there is another view of public reason which, like Rawls’s view, takes as its starting point a fundamental commitment to the freedom and equality of citizens. And this view, like Rawls’s view, takes the commitment to freedom and equality to generate a duty—or requirement—of public justification. The other view to which I refer is convergence liberalism. In this section I contrast convergence *à la* Gaus and Vallier with Rawls’s consensus account. As I will explain below, the difference between consensus and convergence conceptions of public reason is in effect a difference between shared and diverse justificatory reasons.

²⁷ *Id.* at 326.

²⁸ *Id.*

²⁹ Weithman tells us that Rawls’s concern with the assurance problem explains the content of the wide view. (“What citizens know about one another’s commitment to the authority of a conception of justice depends, in part, upon what concepts and methods of reasoning they actually use when they argue about basic political questions. That, I believe, is why Rawls introduces guidelines of public reason—to provide a solution to the assurance problem.”) *Id.* at 159.

3.1 Justificatory Reasons: Shareability and Intelligibility

Justificatory liberalism places a constraint on the permissible use of state coercion or political power. Specifically, it stipulates that political institutions, laws, or basic rules must conform to a principle of public justification. Accordingly, such institutions, laws, or rules are politically or morally justified if and only if each member of the justificatory public—however this public is specified³⁰—has reason to accept them as binding. Call this the *public justification requirement*. In Rawls’s political liberalism, this requirement finds its expression in the “liberal principle of legitimacy” and related “ideal of public reason.” Rawls often discusses both together as a requirement of reciprocity in reason-giving. Rawls tells us that:

[T]he idea of political legitimacy based on the criterion of reciprocity says: Our exercise of political power is proper only when we sincerely believe that the reasons we offer for our political actions—were we to state them as government officials—are sufficient, and we also reasonably think that other citizens might also reasonably accept those reasons.³¹

³⁰ There are, indeed, a variety of ways that Rawls or convergence theorists could characterize the justificatory public—who it consists of and what they know. On David Estlund’s view, for example, membership of the public is restricted to a subset of actual citizens: the justificatory public includes only “qualified points of view”. See, DAVID ESTLUND, *DEMOCRATIC AUTHORITY* 40-64 (2007). On Gaus and Vallier’s view, members of the justificatory constituency are not actual citizens, but actual citizens’ “idealized counterparts,” while for Rawls membership consists entirely of “reasonable” citizens. See GERALD GAUS, *THE ORDER OF PUBLIC REASON: A THEORY OF FREEDOM AND MORALITY IN A DIVERSE AND BOUNDED WORLD* 264 (2011) [hereinafter GAUS, *THE ORDER OF PUBLIC REASON*]. Among the characteristics of reasonable citizens as specified by Rawls is an acknowledgement of the fact of reasonable pluralism and a commitment to satisfying the criterion of reciprocity. Political liberals also tend to idealize members of the justificatory public by assuming they meet a certain standard of cognitive or moral capacity, and that they are free of some of prejudices and biases that plague actual citizens. (For Rawls, the standard is “reasonableness”.) The degree of idealization varies among political liberals. Gaus and Vallier, for example, are critical of what they call the “radical idealization” of Rawls and other political liberals (where choosers are thought to be fully rational, or as in the case of David Gauthier fully rational and fully informed), and instead argue in favor of what they call “moderate idealization,” according to which citizens’ idealized counterparts reason more closely to that of actual citizens. See GERALD GAUS, *THE ORDER OF PUBLIC REASON*, at 276; KEVIN VALLIER, *LIBERAL POLITICS AND PUBLIC FAITH* 145-177 (2014) [hereinafter VALLIER, *PUBLIC FAITH*]. Justificatory liberals also tend to idealize members by assuming they are interested in cooperating with others on mutually acceptable terms. Given these idealizing assumptions, the question political liberals concern themselves with is not with what to do when disagreement emerges among the irrational, prejudiced, uncooperative, etc., but rather what to do when reasonable, good-willed people coming together thinking about politics arrive at different conclusions in their sincere efforts to cooperate.

³¹ JOHN RAWLS, *The Idea of Public Reason Revisited*, in *THE LAW OF PEOPLES* 137 (1999).

There are many versions of the public justification requirement on offer.³² The general idea of Rawls's (consensus) version is that laws and policies are only justified by reasons reasonably acceptable to each citizen,³³ that is, by publicly available (shared) values and standards.³⁴ Under Rawlsian consensus, to show that some proposal is publicly justified, members of the public appeal to reasons they all share in virtue of their common citizenship; they appeal to their shared view of society as a fair system of social cooperation between free and equal citizens and its associated political ideals. When citizens appeal to public reasons, in other words, they appeal to reasons "guided by a political conception the principles and values of which all citizens can endorse."³⁵ Rawls's requirement to publicly justify the exercise of political power follows from (i) the recognition of the "burdens of judgment" and the consequent "fact of reasonable pluralism," and (ii) a commitment to the freedom and equality of all citizens. For Rawls, it is only when the liberal principle of legitimacy is met that the exercise of coercive power is justified. When such a justification obtains, freedom and equality is preserved and respected despite the necessity of political power.³⁶ A more general formulation of the public justification requirement, however, has recently been offered by Vallier and Fred D'Agostino:

³² See, e.g., Paul Billingham, *Convergence Justification Within Political Liberalism: A Defense*, RES PUBLICA 136 (2016) [hereinafter Billingham, *Convergence Justification*]; Gerald Gaus, *Coercion, Ownership, and the Redistributive State*, SOCIAL PHILOSOPHY AND POLICY 53 (2010); Andrew Lister, *Public Reason and Reciprocity*, THE JOURNAL OF POLITICAL PHILOSOPHY 3-4 (2017); VALLIER, PUBLIC FAITH, *supra* note 27, at 24; FRED D'AGOSTINO, FREE PUBLIC REASON: MAKING IT UP AS WE GO 33 (1996); Kevin Vallier, *In Defense of Intelligible Reasons in Public Justification*, THE PHILOSOPHICAL QUARTERLY 8 (2015); JONATHAN QUONG, LIBERALISM WITHOUT PERFECTION 161 (2011).

³³ For an illuminating, in-depth consideration of "reasons all can accept" (RACAs), see James Bohman and Henry Richardson, *Liberalism, Deliberative Democracy, and "Reasons that All Can Accept"*, THE JOURNAL OF POLITICAL PHILOSOPHY (2009).

³⁴ Other consensus theorists include JONATHAN QUONG, LIBERALISM WITHOUT PERFECTION 261-273 (2011); DAVID ESTLUND, DEMOCRATIC AUTHORITY 40-64; Macedo, *Why Public Reason*, *supra* note 6, at 19-35; ANDREW LISTER, PUBLIC REASON AND POLITICAL COMMUNITY 17 (2013).

³⁵ RAWLS, POLITICAL LIBERALISM, *supra* note 8, at 10.

³⁶ *Id.* at 137

PJP: A coercive law L is justified in a public P if and only if each member i of P has sufficient reason(s) R_i to endorse L .³⁷

Public reason liberals like Vallier, Gaus, and D'Agostino locate the main difference between convergence and consensus accounts of public justification in their competing interpretations of the nature of the R variable. On consensus models more generally, R_i is the same for each member of the public. Reason R must be a shared public or political reason (or set of reasons) that each member of the public can accept or endorse.³⁸ Reasoning publicly on this view means reasoning on the basis of shared reasons, as reasons stemming from moral or religious comprehensive doctrines will be the subject of reasonable disagreement. Conversely, Gaus and Vallier's convergence model requires only that reason R be *intelligible* to each member of the public. Gaus and Vallier intend a technical notion of intelligibility. A reason is intelligible when each member of the public can understand and appreciate it as a reason according to the reason giver's own evaluative standards. As a consequence, a publicly justified law L may be based on an intelligible but different justification $R_{i...n}$ for each member of the public i . That is, Alf's religious reason R_A is intelligible—and thus justificatory—when all citizens can see R_A as

³⁷ Kevin Vallier and Fred D'Agostino, *Public Justification*, in STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2018). See also Gerald Gaus and Kevin Vallier, *The Role of Religious Conviction in a Publicly Justified Polity: The Implications of Convergence, Asymmetry and Political Institutions*, PHILOSOPHY AND SOCIAL CRITICISM 53 (2009) [hereinafter Gaus and Vallier, *Religious Conviction*].

³⁸ To say that reason R must be a shared reason is not to say—as some have mistakenly read Rawls as saying—that there may be only one R for which all members of the public accept. Rather, for Rawls, there may be multiple justifications accepted by different citizens, as long as each justification appeals to shared reasons that all citizens accept as reasonable. For further discussion of the distinction between “strong” and “weak” models of consensus, see JONATHAN QUONG, LIBERALISM WITHOUT PERFECTION 263-264 (2011); Jonathan Quong, *Three Disputes About Public Justification: Commentary on Gaus and Vallier* 2 n.2 (unpublished paper, 2008). See also Paul Weithman, *Convergence and Political Autonomy*, PUBLIC AFFAIRS QUARTERLY 338 (2011) (“[F]or Rawls, public reasons are shareable in a relatively weak sense...the fact that public reasons are shareable in this sense does not imply citizens and legislators ‘reason identically.’”). See also RAWLS, POLITICAL LIBERALISM, *supra* note 8, at 241.

justified for Alf according to Alf's reasonable³⁹ religious evaluative standards, even if citizens reject those standards for themselves. Suppose R_A is based on rabbinic legislation enacted for the sake of *tikkun olam*, "repairing the world," that Alf believes is binding. Betty, an atheist, rejects as false those reasons based on rabbinic legislation or any divinely revealed law. Rejecting them for herself, however, Betty still acknowledges rabbinic decree as a reasonable, if false, evaluative standard. On convergence, then, R_A still counts as justificatory, since all citizens (even Betty) can see it as justified according to a reasonable evaluative standard. Consensus, on the other hand, rejects such a reason as lacking justificatory status. Consequently, the intelligibility requirement allows for a broader range of reasons to figure into the justification of a political decision than does Rawls's consensus standard, since intelligible reasons can figure into the justification of a given law despite those reasons being unacceptable according to others' evaluative standards. In this way, members may *converge* on a given law for wholly different reasons.⁴⁰

Convergence theorists advance two main arguments in favor of convergence over more familiar consensus models of public justification, including Rawls's. The first, which I will not pursue here, is that convergence better respects reasonable pluralism and places fewer restraints on individual liberty than does consensus.⁴¹ Convergence theorists wonder, if the purpose of public justification is to ensure that political power is only exercised over someone if that power

³⁹ Here Gaus and Vallier do not mean "reasonable" in the sense specified by Rawls.

⁴⁰ Gaus and Vallier argue that there is an asymmetry "between reasons to *justify to another* a law and reasons to *reject* that law." Gaus and Vallier, *The Role of Religious Conviction*, *supra* note 36, at 52 (emphasis in original). This implies that reasons for supporting a proposal, on the one hand, and reasons for rejecting that proposal, on the other, are not subject to the same requirements. Concerning religious convictions, for example, the idea is that it would be wrong to impose a law L solely on the basis of religious reasons since some members of the public will not hold within their belief-set such reasons, but religious reasons can still justifiably *defeat* L .

⁴¹ VALLIER, PUBLIC FAITH, *supra* note 27, at 45-72; Kevin Vallier, *Convergence and Consensus in Public Reason*, PUBLIC AFFAIRS QUARTERLY 4-6 (2011) [hereinafter Vallier, *Convergence*].

can be justified to her, why demand that the reasons justifying the power in question be shared reasons? If a law *L* can be justified to each member of the public by a convergence of different reasons, so the thought goes, such an agreement could satisfy the requirement to publicly justify coercive power just as well as consensus, while at the same time not placing unfair burdens on persons of faith by excluding their reasons as justificatory.⁴² The second is that convergence offers greater stability than does consensus. Specifically, convergence discourse's signaling device serves as a better assurance mechanism than does consensus's signaling device.

4 CONVERGENCE DISCOURSE AND THE PROBLEM OF MUTUAL ASSURANCE

Recall the problem of mutual assurance. Citizens with a sense of justice prefer to comply with just principles, even against their personal interests. However, even with a sense of justice they are unwilling to make such sacrifices unilaterally. Rawls remarks, "The sense of justice leads us to promote just schemes and to do our share in them *when we believe that others, or sufficiently many of them, will do theirs.*"⁴³ Thus each citizen who is inclined to comply with the demands of justice needs to be assured that others will do the same. Without assurance of others' compliance, a citizen will reasonably worry that others will take advantage of her just behavior. Merely having an effective sense of justice, then, is not enough to secure the stability of a society well-ordered by a reasonable political conception of justice. The fact that each has an effective sense of justice must also be common knowledge. As stated above, on Weithman, Macedo and

⁴² Billingham, *Convergence Justification*, *supra* note 29, at 136-139; Christopher Eberle, *Consensus, Convergence, and Religiously Justified Coercion*, PUBLIC AFFAIRS QUARTERLY 285 (2011); Vallier, *Convergence*, *supra* note 39, at 2-6; Gaus and Vallier, *The Role of Religious Conviction*, *supra* note 34, at 53-56; Gerald Gaus, *The Place of Religious Belief in Public Reason Liberalism*, in MULTICULTURALISM AND MORAL CONFLICT (Maria Cookson and Peter Stirk eds., 2009); GAUS, THE ORDER OF PUBLIC REASON, *supra* note 27, at 276-292.

⁴³ RAWLS, THEORY, *supra* note 3, at 236 (emphasis added).

Hadfield's reading of later Rawls, public reason solves this problem. In recent years, however, a number of convergence theorists have for various reasons come to doubt the possibility of Rawls's public reason solving the assurance problem. Gaus,⁴⁴ and John Thrasher and Vallier⁴⁵ argue that Rawls's public reason cannot solve the assurance problem because public reason is mere *cheap talk*. Brian Kogelmann and Stephen G. W. Stich⁴⁶ agree that public reason is *cheap talk*, but argue that the *cheap talk problem* can be overcome by resorting to convergence discourse as *costly signals*, thereby solving the assurance problem.

Thus much of the current convergence scholarship focuses on various signaling devices—direct and indirect—that serve as assurance mechanisms. But in focusing on competing signaling devices, they focus on one component of a much larger argument for generating and maintaining the mutual assurance of members of a well-ordered society. In the following section, I outline the dominant criticisms leveled against consensus public reason as an assurance mechanism in the existing literature and evaluate the alternative mechanisms that convergence theorists have proposed. Along the way I gesture towards some of the advantages of consensus over convergence, though I do not claim these remarks constitute a positive argument in favor of consensus. Rather, I suggest we have less reason to be optimistic about the stability that convergence offers than convergence theorists recommend. One question that is likely to arise throughout my discussion is whether the real-world problems that I claim threaten convergence as an assurance mechanism are relevant to Rawls's well-ordered society. I submit

⁴⁴ Gerald Gaus, *A Tale of Two Sets: Public Reason in Equilibrium*, PUBLIC AFFAIRS QUARTERLY 317 (2011) [hereinafter Gaus, *Tale of Two Sets*].

⁴⁵ John Thrasher and Kevin Vallier, *The Fragility of Consensus: Public Reason, Diversity and Stability*, EUROPEAN JOURNAL OF PHILOSOPHY 943-946 (2013) [hereinafter Thrasher and Vallier, *The Fragility of Consensus*].

⁴⁶ Brian Kogelmann and Stephen G. W. Stich, *When Public Reason Fails Us: Convergence Discourse as Blood Oath*, AMERICAN POLITICAL SCIENCE REVIEW 5 (2016) [hereinafter Kogelmann and Stich, *When Public Reason Fails Us*].

they are. The problems I discuss are problems that arise even among members understood as rational and reasonable, motivated by reciprocity.

5 THE FRAGILITY OF CONVERGENCE

The main appeal of the consensus view is found in its attention to reasonableness, reciprocity, and respect. But as much of my discussion has tried to bring out, a further need which convergence theorists must attend to is political stability. This section distinguishes between what I shall call *strong* and *weak* stability. *Strong* stability obtains in a society when its basic institutions, taken together, generate their own support.⁴⁷ Recall that basic institutions generate their own support in two ways. First, institutions generate their own support by eliciting a sense of justice. Second, institutions bring it about that citizens judge it rational to maintain their sense of justice rather than defect. Institutions are able to support themselves in the second way only if the assurance problem is overcome. *Weak* stability, on the other hand, obtains in a society when its basic institutions tend to generate their own support, but they do not do well at correcting themselves when infractions occur. Moreover, *weak* stability is stability that is likely to persist for periods of time but is liable to disruptions. Consensus and convergence theorists alike are interested in *strong*, rather than *weak*, stability.

On one side we have Thrasher and Vallier's convergence model of indirect public reason, which I argue represents a form of *weak* stability. On the other side is Kogelmann and Stich's convergence model of costly signals, which I argue represents a form of *strong* stability. Nevertheless, for reasons that will become clearer, neither is an advance over the stability that Rawlsian shared reasoning offers.

⁴⁷ In this I follow Rawls's "inherent" stability. *See supra* note 11.

In criticizing Rawls's solution to the assurance problem, convergence theorists take as their starting point the following. Critics charge that when Alf and Betty engage in public reason, Alf cannot infer from Betty's adherence to its guidelines that Betty will actually choose to maintain her sense of justice rather than defect.⁴⁸ In other words, Betty's public display of constrained deliberation does not provide Alf assurance that she will actually maintain her sense of justice. Recall that the assurance problem arises on the assumption that members already have a sense of justice—due to the forces of social learning at work in a well-ordered society—but need to be reasonably sure that others are going to continue to *maintain* their sense of justice if they are to maintain their own. Implicit in convergence theorists' worry, then, is that adhering to the guidelines of public reason is *cheap talk*: “We can understand Rawlsian displays of shared public reasoning as what economists call ‘cheap talk’.”⁴⁹ Thrasher and Vallier define cheap talk as “costless or very inexpensive, non-binding communication in a game.”⁵⁰

Thrasher and Vallier develop one way of understanding the worry that adhering to the guidelines of public reason is cheap talk. On their view, the problem with using Rawlsian “direct public reason”—whereby citizens directly assure one another of their commitment—as a means

⁴⁸ Gaus, *Tale of Two Sets*, *supra* note 40, at 317-18; Kogelmann and Stich, *When Public Reason Fails Us*, *supra* note 42, at 5. Thrasher and Vallier, *The Fragility of Consensus*, *supra* note 41, at 943-946.

⁴⁹ Gaus, *Tale of Two Sets*, *supra* note 40, at 317. The name “cheap talk”, however, is a little misleading. The claim is not necessarily that talk is cheap, but that it is too cheap—or not costly enough—to ensure that it is rational for members to adhere to the guidelines of public reason *only* if they plan to maintain their sense of justice. In other words, adhering to the guidelines of public reason is too cheap to render one's adherence to such guidelines rational only if she plans to maintain her sense of justice. (“Alf's communication that he intends to cooperate will not help Betty form beliefs about what Alf will do, because regardless of what Alf intends he would have a reason to send an ‘I will cooperate’ message. If Alf plans to cooperate he would tell Betty so (since he will get his best payoff if she also cooperates), but if he plans to defect he *still* has reason to give a display of trustworthiness, since he prefers his unilateral defection to mutual defection, and so has an incentive to induce Betty to cooperate.”) *Id.* at 318.

⁵⁰ Thrasher and Vallier, *The Fragility of Consensus*, *supra* note 41, at 943.

to political stability is that it is open to “noise” and “amplification”. The wide view of public reason creates a problem of noise because once citizens are allowed to introduce reasons stemming from their comprehensive doctrines, it will be difficult to distinguish reasons based on the political conception from those that are not so based: The problem becomes one of distinguishing between communication that signals allegiance to the political conception, thus providing assurance, and communication that does not.⁵¹ Since the public provision of reasons is the main way that citizens signal their allegiance to the political conception, Thrasher and Vallier tell us, true signals of allegiance must be easily distinguishable from noise.⁵² The second problem for direct public reason is the amplification of noise: small errors in communication, multiplied over large numbers of interactions, “can create ‘informational cascades’ that can dramatically undermine mutual assurance consensus equilibria.”⁵³

In response to these alleged problems with Rawls’s conception of public reason, Thrasher and Vallier develop an alternative account on which stability owes to a process of “indirect public reason”, that is, “without direct, deliberative assurance that all will affirm the political conception.”⁵⁴ On this view, assurance is generated and maintained through social mechanisms known as “choreographers”. For Thrasher and Vallier, these “public choreographers are primarily bodies of norms, often legal, though sometimes informal or formal moral norms.”⁵⁵ As

⁵¹ *Id.* at 942. One way to avoid this problem is to move from the wide view to the exclusive view, in which case, when fundamental political matters are at stake, comprehensive reasons are never to be introduced into public reason. Thrasher and Vallier admit that a “nice feature of the exclusive view is that it avoids [the problem of] noise,” but insist that “[s]uch restrictions should be unattractive to political liberals who care about protecting freedom of speech and expression.” *Id.* at 946.

⁵² *Id.*

⁵³ *Id.* at 944. The general worry is that, much in the same way that in stock markets financial harm is brought upon when the initial source in an information cascade is unreliable, so too is mistaking noise for assurance-generating communication harmful to assurance. *Id.* at 944-945.

⁵⁴ *Id.* at 934.

⁵⁵ *Id.* at 948.

long as citizens follow the relevant public choreographers, and know that others do so too, the assurance problem is overcome and the political conception is stable. Their running example is a traffic light at a high-traffic intersection. Thrasher and Vallier ask us to imagine the following scenario. First, imagine a case with no stop signs or lights. Assurance mechanisms are needed in order for drivers to safely cross the intersection without colliding into one another. Absent signs and lights, drivers must directly assure one another by coordinating with their own lights, horns, and eye contact. “This,” Thrasher and Vallier tell us, “is directly analogous to the conception of direct public reason.”⁵⁶ While we can improve matters with four-way stop signs, as the number of vehicles increase, the amount of direct coordination needed becomes more costly since drivers must still determine who first arrived at the stop. Their “obvious solution” to the traffic problem is to install a traffic light. “Traffic lights correlate coordination among drivers to an independent, public signal.”⁵⁷ By following a traffic light, Thrasher and Vallier tell us, “drivers no longer need to directly assure one another of their intentions by signaling”; rather, “the choreographer generates assurance indirectly.”⁵⁸ The traffic light forms and maintains assurance of others’ compliance and citizens no longer need to coordinate directly with one another. As the traffic light example is meant to bring out, for publicly recognized choreographers, citizens will simply “realize that it is in [their] best interest to follow the public signal.”⁵⁹

As I understand it, three problems face the Thrasher-Vallier model of indirect public reason through publicly recognized choreographers. I only mention the first two and focus on

⁵⁶ *Id.* at 946

⁵⁷ *Id.*

⁵⁸ *Id.*

⁵⁹ *Id.*

the third. First, the task remains of identifying relevant choreographers capable of providing assurance.⁶⁰ Second, a system of indirect public reason pushes deliberation to the periphery. At the very least, public discussion plays a less central role than most political liberals might wish.⁶¹

But the main difficulty with the use of indirect public reason, particularly as it pertains to assurance, is that it relies on citizens deciding to be just issue-by-issue. Specifically, Alf's obeying traffic laws at an intersection will not assure Betty that he will serve on a jury when called upon, or that he will not cheat on his taxes. On an indirect model, Alf can only provide assurance to Betty that he will not act on his temptation to, say, desert his post in times of war if Alf, in fact, does not desert his post (or does not defect in a suitably similar situation). On the Thrasher-Vallier model, members are left to factor the payoffs associated with the good of not maintaining their sense of justice on an issue-by-issue basis. Perhaps in some instances members will choose to maintain their sense of justice even when strong reasons tell against acting justly, and perhaps in some instances members will maintain their sense of justice even when it is no longer in their best interest to do so. But in other cases, members may not. The main problem

⁶⁰ Though I do not have the space here to defend, I think this difficulty can be overcome, though its solution departs from the type of stability Rawls envisioned.

⁶¹ Specifically, Thrasher and Vallier, and here they join Gaus, reject the importance of deliberation for public justification: what matters is not that citizens or public officials participate in a democracy that is deliberative, but rather that coercive laws are *justified* from the point of view of each citizen. See GERALD GAUS, *JUSTIFICATORY LIBERALISM: AN ESSAY ON EPISTEMOLOGY AND POLITICAL THEORY* 10-12 (1996); Gaus and Vallier, *The Role of Religious Conviction*, *supra* note 36, at 65-70. The Thrasher-Vallier model of indirect public reason is emblematic of this view. But this view, whatever its merits, comes at a cost. For example, to the extent that they are carried out on fair terms, part of the importance of deliberative political practices for a healthy and stable democracy are their contribution to fostering a common—civic—identity, a sense among members that they are equal participants in a common enterprise with a shared responsibility for the fate of their collective political project. See also JOHN RAWLS, *JUSTICE AS FAIRNESS: A RESTATEMENT* 40, 93-94, 118 (Erin Kelly ed., 2001) [hereinafter RAWLS, RESTATEMENT]. See also Blain Neufeld's "civic people" account of public reason. Blain Neufeld, *Shared Intentions, Public Reason, and Political Autonomy*, *CANADIAN JOURNAL OF PHILOSOPHY* (2018). See also RICHARD DAGGER, *Citizenship as Fairness: John Rawls's Conception of Civic Virtue*, in *A COMPANION TO RAWLS* 297-311 (Jon Mandle and David Reidy eds., 2014). Even if we reject deliberative political practices as important for public justification, it would be a mistake to not encourage their use given their other crucial roles.

with members deciding to be just issue-by-issue, particularly as it pertains to assurance, is that it will be difficult for members to determine whether the relevant others—or a sufficient many of them—will actually follow the appropriate choreographer or shirk. In response, perhaps convergence theorists can suggest citizens simply look to current desertion rates and factor their fellow citizens' chances of defection. As long as they can predict that a sufficient many will follow the relevant norm, then all is well. This suggestion, however, opens assurance up to further questions and complications not found under consensus, questions and complications about which citizens may reasonably come to doubt. Rawls sought to keep the principles of justice and the guidelines of public inquiry as simple and straightforward as possible to “insure that public reasoning can publicly be seen to be—as it should be—correct and reasonably reliable in its own terms ... Hence applying liberal principles has a certain simplicity.”⁶² Further complicating the process of securing assurance does nothing in the way of putting doubts to rest. Moreover, mere observation of others' compliance without knowledge of their rationale is open to wide interpretation. Citizens may reasonably wonder whether others' reasons for complying are good reasons—that is, stable reasons—or capricious, etc., ones. This problem is only further compounded the more diverse a society becomes.

Abandoning the publicity of rationales renders stability under Thrasher and Vallier's model *weak* for the following reason. A society's basic institutions under the Thrasher-Vallier model will tend to generate their own support since, after all, citizens will recognize that it is in their best interest to follow certain public choreographers, such as traffic lights. But this is not always so for other public signals, and citizens may reasonably worry whether a sufficient many will, say, cheat on their taxes if they think they can get away with it. Citizens need not believe

⁶² RAWLS, POLITICAL LIBERALISM, *supra* note 8, at 162.

that it is likely that their fellow citizens will cheat on their taxes in order for assurance to be compromised; citizens need only come to doubt in order for defection to become the dominant strategy. Recall that Rawls thought members of the well-ordered society have reason to defect if they come even to doubt—and are not reasonably assured—that others will do their part. Rawls saw assurance as something that builds over time.⁶³ The type of assurance Rawls envisioned entails a commitment to forgo the gains of injustice, come what may, rather than leaving oneself open to choose the action promising them the greatest return. Rawls remarks, “we cannot preserve our sense of justice and all that this implies while at the same time holding ourselves ready to act unjustly should doing so promise some personal advantage.”⁶⁴ For Rawls, members of the well-ordered society are not motivated to follow traffic laws and to pay their taxes because failure to do so violates the norms of each, or because they see compliance as in their best interest. Rather, they do so because they have a stronger preference for mutual cooperation—and for realizing other ideals of free and equal citizenship—that normally outweighs the gains of breaking traffic laws and evading taxes. That the gains of injustice are normally outweighed in this way is one of the ways that institutions informed by the political conception generate their own support.

I thus characterize Thrasher and Vallier’s model as representing a form of *weak* stability: while basic institutions will tend to generate their own support due to citizens following, I shall grant, *most* publicly recognized choreographers, sufficient doubts may arise over whether their fellow citizens will follow all other choreographers. This situation becomes especially problematic when the “inevitable” injustices that Rawls speaks of occur: Basic institutions will

⁶³ Rawls states this more clearly later in RAWLS, RESTATEMENT, *supra* note 60, at 197.

⁶⁴ RAWLS, THEORY, *supra* note 3, at 498. Rawls goes on to say that “the just person is not prepared to do certain things.” *Id.*

not do well at restoring themselves when infractions occur if restoration comes at too great a cost relative to citizens' amount of assurance. Absent right reasons, citizens lack sufficient assurance to willingly do their part to restore institutions when injustices occur. Injustices, Rawls tells us, are only a matter of time, even in a well-ordered society.⁶⁵ *Strong* stability requires a willingness on the part of citizens to restore institutions when these inevitable injustices occur. But when citizens' compliance with choreographers owes to their "realiz[ing] it is in [their] best interest"⁶⁶ to do so, a willingness to restore institutions is on less solid grounds. Stability, in other words, is fragile. To the extent that indirect public reason avoids the problems of "noise" and "amplification", it does so at the expense of *strong* stability. Note that none of this requires that citizens be unreasonable or no longer motivated by a desire to answer in kind; these are problems that occur even in the well-ordered society.

Kogelmann and Stich offer a different solution to the *cheap talk problem*. They argue that the cheap talk problem can be overcome by resorting to convergence discourse as costly signals: Successful participation in convergence discourse requires a considerable network of communication and knowledge about a wide range of comprehensive doctrines other than one's own.⁶⁷ The burden of learning the relevant doctrines, according to Kogelmann and Stich, furnishes convergence discourse with a costly enough assurance mechanism to overcome the cheap talk problem. Let us suppose that Alf, whom Betty does not anticipate knows much about Jainism, offers an elaborate Jain-based argument. Since Alf is not a Jain, Betty knows that Alf

⁶⁵ RAWLS, THEORY, *supra* note 3, at 401.

⁶⁶ Thrasher and Vallier, *The Fragility of Consensus*, *supra* note 41, at 946. Under the Thrasher-Vallier model, compliance mechanisms appear to function more like a market than agreement based reciprocity, though I do not have the space here to defend this claim.

⁶⁷ Indeed, convergence's considerable informational burdens is precisely why Blain Neufeld remarks that convergence "seems unfeasible." See Blain Neufeld, *Shared Intentions, Public Reason, and Political Autonomy*, CANADIAN JOURNAL OF PHILOSOPHY 9-10 (2018).

put considerable effort into learning the doctrine well enough to offer his argument. Betty knows that Alf incurs opportunity costs when he makes the decision to learn the relevant doctrine. That Alf is willing to incur such costs signals he is serious about achieving the mutual assurance equilibrium. At the heart of convergence as costly signals, Kogelmann and Stich tell us, is that members “will have to engage in what Rawls calls *reasoning by conjecture*,” whereby “we argue from what we believe, or conjecture, are other people’s basic doctrines, religious or secular, and try to show them that, despite what they might think, they can still endorse a reasonable political conception.”⁶⁸

An interesting corollary of Kogelmann and Stich’s view, as they explain it, is that the more diverse a society is the better its ability to overcome the assurance problem: convergence discourse becomes costlier the more diverse a society becomes. This seems obviously correct. Signaling costs can only increase with diversity; signals impose a cost on the signaler that can only increase as diversity increases.⁶⁹ To the extent that there was ever a cheap talk problem, I think Kogelmann and Stich’s convergence model has the resources to overcome it. If convergence discourse as costly signals could be realized, it would be sufficiently costly to

⁶⁸ There is, however, a difference between Rawls and Kogelmann and Stich’s uses of reasoning from conjecture. For Rawls, the purpose of reasoning from conjecture is to try to reconcile conflicts between private and public values: reasoning from conjecture is Alf’s attempt to show Betty, a Jain, how she might reconcile her faith in Jainism with a moral commitment to the political conception. For Kogelmann and Stich, convergence discourse demands reasoning from conjecture as a means to “(1) successfully convince others to support their respective position and to (2) also engage with [them] on terms [they] can endorse.” Kogelmann and Stich, *When Public Reason Fails Us*, *supra* note 44, at 9. For a rich defense of how reasoning from conjecture can be respectful and sincere, see Micah Schwartzman, *The Ethics of Reasoning from Conjecture*, *JOURNAL OF MORAL PHILOSOPHY* 529-33 (2012). *But cf.* QUONG, *LIBERALISM WITHOUT PERFECTION* 265-75 (2011).

⁶⁹ Moreover, signaling costs would seem proportionately greater with more diversity, though this is not a claim that Kogelmann and Stich make. I think a great virtue of Kogelmann and Stich’s view is that, as costly as signaling may become, the cost is negligible relative to the benefits of signaling. Hence we could expect a willingness on the part of citizens to continue trying to signal despite the increased cost of signaling.

secure *strong* stability. Nevertheless, it is impossible if not at least highly impractical to have a working model of costly signals in diverse societies: convergence as costly signals underestimates the difficulty of constructing and maintaining ongoing political arrangements in societies marked by deep pluralism and pervasive disagreement, even where those societies are well-ordered. Communicators may in fact be sincere in their signaling, but if assurance depends upon citizens being able to publicly justify laws in terms of all their fellow citizens' diverse doctrines, and to engage issues on the public agenda in the language of all those diverse doctrines, such assurance would hopelessly complicate the deliberation and implementation processes. There is not only great moral value, but great administrative value in having a shared language and rationale that everyone, for all their differences, can readily access.⁷⁰ Attempts to discuss laws and policies in terms of every possible comprehensive doctrine, rather than approaching the public agenda with a shared sense of the relevant considerations, renders the deliberative and administrative processes even more intractable than they are under the already complicated conditions of shared reasoning. I do not claim that convergence under Kogelmann and Stich's model is a form of *weak* stability so much as that it cannot be realized in diverse societies. The mechanism that makes assurance, I will grant, sufficiently costly is the very mechanism that makes a working model of it, for all practical purposes, unrealizable.

6 CONCLUDING REMARKS

The attraction of an account of public justification will depend in part on its attention to our deepest moral convictions regarding reasonableness, reciprocity, and mutual respect. I think these convictions receive their best expression in a consensus, shared reasons account. But its

⁷⁰ For a discussion of the value of a shared rationale for interpreting and administering newly enacted laws, and extending them to new circumstances, see Macedo, *Why Public Reason*, *supra* note 6, at 22-24.

attraction will also depend on its ability to maintain a just equilibrium over time. I have suggested that convergence as presented by Thrasher and Vallier, on the one hand, and Kogelmann and Stich, on the other, is not better suited to this task than consensus. At the very least, I have offered reasons to suggest we be less optimistic about the stability convergence offers than convergence theorists appear to be.