American Political Science Review (2018) 112, 2, 423–427
doi:10.1017/S000305541700065X © American Political Science Association 2018

Letter

Convergence's Democracy Problem: A Critique of Kogelmann and Stich

BENJAMIN R. HERTZBERG *Emory University*

In their 2016 American Political Science Review article, Kogelmann and Stich argue that public reason fails to provide the assurance reasonable citizens require to act justly and that, as a consequence, Rawls's account of political stability fails. Convergence discussions, because they are a costly signal, provide such assurance. Kogelmann and Stich fail to recognize that constituents influence representatives such that the costs of convergence discourse are unknown. It thus cannot assure. Constituents' influence also undermines convergence's ability to show how decision-making processes that follow its norms result in justified laws. Far from supporting convergence, then, the stability question demands revision of the view. This response develops these objections, extends them from Kogelmann and Stich's analysis to other convergence theorists and political liberals, and explores what political theorists can learn from convergence's difficulties.

How should democratic citizens create enduring relationships of civic trust in a context of deep moral and religious pluralism? Political liberals refer to this as the question of "stability." In Kogelmann and Stich's article, "When Public Reason Fails Us: Convergence as Blood Oath," the authors develop a new account of stability (2016). They argue that the dominant account, based on Rawls's ideal of public reason, fails, but one based on the "convergence" conception of public justification succeeds. I believe they are wrong: the stability question demands revision of convergence as well as other recently popular political liberal positions. Both positions unjustifiably ignore the democratic context of the stability question.

"Convergence" denotes the claim that laws are publicly justified when each citizen has a conclusive reason for supporting the law. This contrasts to Rawls's claim that only public reasons—reasons that all citizens can reasonably be expected to endorse—justify law (2005, 450–8). Kogelmann and Stich (hereafter K&S) argue that convergence encourages forms of public discussion that better provide the assurance reasonable citizens require to act justly. Philosophers have subjected the convergence conception to considerable analysis (Eberle 2011; Gaus 2011b; Lister 2011; Vallier 2011; Weithman 2011; Gaus 2011a; Gaus and Vallier 2009; Gaus 2010; Thrasher and Vallier 2015; Vallier 2014). K&S are the first to advocate for it in a political science journal and to explore its implications for stability. As developments in the Rawls literature demand (Hadfield and Macedo 2012; Weithman 2010), K&S's analysis is game theoretic. This results in a formal account that may appeal to many political scientists.

However, K&S fail to consider two crucial issues that undermine their account. First is the influence of citizens' activism on the legislative process. This influence prevents public officials from accurately assessing the costs of convergence discourse, costs that are essential to its ability to provide assurance. This leads to the second issue: K&S fail to recognize that assurance comes crucially from what people do and not only from what they say. These two issues together demand revisions to the convergence conception generally, not only K&S's account of it. In the context of a representative relationship between citizens and public officials, convergence does not offer a plausible account of how convergence-justified laws result from the kinds of political action the conception permits. And if political decision-making processes do not create publicly justified laws, reasonable citizens cannot be expected to endorse those laws or trust the system that produces them. For these reasons, convergence has a democracy problem, one grounded in institutional questions and distinct from the objections of democratically inclined philosophers (Weithman 2011; Watson 2015). Below, I develop my criticisms of K&S's claims, extend my analysis to other discussions of convergence, and explore what political theorists can learn from convergence's difficulties.

K&S make three essential claims. They begin with a newly popular position among political liberals, that stability requires solving an assurance game (Hadfield and Macedo 2012; Weithman 2010). As K&S claim, "those with a sense of justice aren't suckers" (717). This corresponds to Rawls's assumption that citizens behave reciprocally; they treat people justly if they can be assured others will treat them justly in turn (Rawls 2001, 16–8). The key question is how to generate such assurance. K&S argue that political liberals have ignored a "puzzle" in Rawls's discussion of stability. K&S believe that political liberals think citizens assure each other of their commitment to just action by offering public reasons to support the laws they propose. But K&S also (inaccurately) believe that Rawls limited public reasoning requirements to public officials and did not apply them to citizens. K&S claim that political liberals must explain how public officials' discourse will

Benjamin R. Hertzberg is a visiting Assistant Professor of Political Science, Emory University (bhertzberg@emory.edu).

I am grateful to Ryan Davis and the APSR's anonymous reviewers for their feedback on earlier versions.

Received: June 20, 2017; revised: October 6, 2017; accepted: December 5, 2017.

https://doi.org/10.1017/S000305541700065X Published online by Cambridge University Press

assure citizens of their fellows' intentions to act justly. Just because the local mayor offers public reasons does not mean that my neighbor will treat me justly (719).

K&S's solution to this apparent puzzle is to argue that Rawls divided the assurance problem in two: public officials face a different assurance problem than normal citizens do. Public officials have to determine whether to trust their colleagues: acting justly for public officials means justifying proposals in terms of public reasons and passing laws that public reasons support. But citizens are different: "because citizens do not have the capacity to legislate," K&S claim, acting justly for them means "that citizens simply obey laws" (719). So, normal citizens merely need to know that their fellows will obey the laws to be assured. The standard coercive apparatus of the state provides such assurance.

K&S have badly misconstrued Rawls here. Rawls explicitly claimed that public reason is "the reason" of democratic citizens generally and requires all citizens, public officials or not, to follow its dictates when they participate in politics (2005, 213, 214, 217–8). Furthermore, assurance, on Rawls's account, comes not merely from public officials' discourse but rather from citizens' knowledge that there is an overlapping consensus (2005, 133–72). They know—and realize that their fellows know—that everyone is committed to a liberal, political conception of justice, knowledge that arises as much from being treated justly as they were raised within a just society as from the verbal assurances that citizens or public officials offer each other in political debate (Rawls 1999, 397–449). Analyzing K&S's position is worthwhile despite this misreading because many political liberals, advocates of convergence and otherwise, excuse normal citizens from the obligations of public reason. K&S's position, properly understood, shows why this undermines rather than enhances political liberal stability.

K&S's second key claim is that public reason (as K&S understand Rawls's conception) fails to offer public officials sufficient assurance of their colleagues' intentions to act justly. Public reasoning is "cheap talk" (720–1). In game theory, cheap talk is communication that neither affects the payoffs of the players nor reliably communicates players' intentions. It is an endemic issue in many games because players do better if their colleagues act justly while they do otherwise. Players have an incentive to dissemble. Wise players know this, so they do not take others' communication as trustworthy. Public reasoning, for K&S, falls into this problem: it is easy for a public official to offer public reasons falsely, claiming that they justify a law she knows to be unjust (720–1). So public reasoning does not provide assurance.

K&S's final claim is that convergence encourages elite political discourse that provides assurance to public officials. "Convergence discourse," they claim "is a costly signal" (724). Advocates of convergence claim that if all citizens have a conclusive reason for a law, the law is justified, even if the reason arises from citizens' comprehensive doctrines. K&S argue this conception leads to political deliberations in which, instead of searching for public reasons to support proposals, officials search for reasons that arise from their colleagues' comprehensive doctrines. In a discussion between a Christian representative and a Hindu one, K&S imagine the Christian can give "Hindu-based reasons" (725) to her colleague. Doing so requires that she "learn" her colleague's "comprehensive doctrine to a significant degree" (ibid.). The time K&S's Christian invests in learning Hindu reasons and using them to persuade her colleague is costly; it thereby signals to him her intention to act justly. Convergence discourse is thus, for K&S, like a blood oath (723): the search for reasons to support one's preferred laws from within a comprehensive doctrine that one rejects is like cutting open one's hand to demonstrate commitment. Such costly actions are the opposite of cheap talk; they reliably demonstrate people's intentions to act justly. On this basis, K&S believe they have "succeeded" in solving the public official assurance problem while Rawls "fails" (729).

However, K&S have not given a persuasive account of how convergence discourse provides assurance. Blood oaths generate assurance only when given two crucial conditions apply: first, the cost incurred by the oath must be publicly recognized (as cutting one's hand is), and second, that cost must be sufficiently high to outweigh whatever utility the oath taker could gain by using it to trick other players. (These are the explicit terms of K&S's discussion and models, see 723–4.) Players must have a relatively clear understanding of other players' incentives and values (utility functions) for blood oaths to operate in the way K&S describe. But it is not clear that K&S's public officials have the kind of knowledge of their colleague's utility functions that K&S's account requires.

K&S miss this difficulty because of the strong distinction they draw between public officials (who legislate) and citizens (who do not). This distinction gives false precision to K&S's models;[1] it allows them to formally ignore constituents' influence on representatives. K&S are of course correct that ordinary citizens do not directly propose and pass laws in legislatures. But they do participate in legislation in many other ways: voting, protesting, calling representatives, donating to campaigns, arguing in the media, and so on. And such participation is dynamic: it changes as the legislative agenda changes, as different groups mobilize, or contribute, or withhold funds, and so on. Such participation deeply shapes representatives' utility functions; for representatives require citizens' support not only for reelection but also in support of their legislative initiatives. Thus actions that were beneficial yesterday could quite quickly become costly today, and vice versa, and the events that underlie such changes may depend on highly local (not publicly known) circumstances: a key donor changes her position; a crucial civic group makes a credible, private threat to withdraw support; and so on. Such dynamism in a representative's utility function means that convergence discourse will often not meet either of the conditions required for assurance: its cost

[1] My thanks to an anonymous reviewer for this language.

https://doi.org/10.1017/S000305541700065X Published online by Cambridge University Press

may not be sufficiently high or publicly known that it is sufficiently high.[2]

Once representatives are considered in light of their connections to constituents, it's clear that K&S have not considered the factors that would lead to an accurate assessment of the costs of convergence discourse. They almost exclusively focus on the costs of learning other comprehensive doctrines; they only discuss constituents' responses to their representative's convergence discourse in a footnote (726n6). K&S do correctly note that conservative Christian constituents may impose costs on Rick Santorum-like representatives who publicly make Hindu or Muslim arguments. But this is highly contextual. Perhaps the Christian and Hindu share substantive commitments about laws regulating human sexuality, and so when the Christian offers Hindu reasons for such laws, her conservative constituents see it as strategic coalition-building and do not punish her. Perhaps she has Hindu constituents whose support she will win by making Hindu arguments. Whether convergence discourse is costly to the Christian representative, then, depends only superficially on the time she invests in learning Hindu arguments: the real action is in the political consequences of her so speaking on the dynamic relationship between her and her constituents. But these are difficult to predict: American representatives invest large sums in opinion polling and focus groups and still regularly misjudge their constituents. So it is far too quick to assume, as K&S's position requires, that the Hindu representative will recognize Hindu arguments from his Christian colleague as a self-imposed cost akin to a blood oath. And this is generalizable: given the dynamic relationship between representatives and constituents, representatives are consistently unable to assess the costs convergence discourse imposes on their colleagues. If the costs of convergence discourse are dynamic and not publicly known, then convergence discourse cannot provide assurance. It is not a blood oath.

K&S realize that their readers will worry that convergence discourse is not costly enough to provide assurance, but this is a different concern. In their telling, the worry is that public officials will transfer the costs of learning their colleague's comprehensive doctrines to their staff members. Their response is that, in legislative deliberations, representatives will need to be able to think quickly within the terms of their colleague's comprehensive doctrines, and this will force them to spend the time necessary to learn those doctrines and to keep their knowledge of them up-to-date (727–8; 725–7). The deeper objection, however, is that K&S's unrealistic formal separation of representatives from constituents allows them to assume that representatives' utility functions are public to their colleagues in a way that does not correspond to actual politics. The problem is not that convergence discourse is not costly enough; it is rather that representatives cannot confidently assess its costs.

The second problem in K&S's argument is also a product of their inaccurate assumption that representatives operate largely without influence from their constituents. K&S are the first to systematically analyze convergence's implications for political liberal stability. Against K&S, however, this question leads to a problem for convergence, one with broader implications for some traditional political liberals who hold relevantly similar views.

To see K&S's problem, it is crucial to recognize that assurance cannot arise from discourse alone. Discourse is too easily feigned; it is too often cheap talk. K&S's article is built on this assumption: blood oaths are valuable because they are publicly recognized as costly actions that, given the parameters of certain games, communicate sincerity. In iterated games, there are additional ways for players to signal their trustworthiness: namely, by acting justly or cooperatively in previous interactions. This is why tit-for-tat strategies are effective in many iterated games: once players know that their fellows are cooperators, they can confidently cooperate too (Binmore 2005, 77–92). Rawls recognized this; it is a deep reason why all of his various accounts of stability are developmental: he took game theory, as well as deeper claims about human psychology, to show that relationships of reciprocal trust arise in part from treating others justly over time (2005, 133–72; 1999, 397–449; Weithman 2010, 341–3).[3] It is thus crucial for a theory of public justification to show that the implied political decision-making process is likely to produce decisions that meet its criteria of justification. Otherwise, it will not generate stability because stability requires that people be treated justly, which in turn requires that enacted laws are properly justified (by whatever criteria the theory advocates).

As a consequence of the strong distinction K&S draw between the norms that apply to constituents' actions and those that apply to public officials, they cannot give a plausible account of how the political decision-making processes theorists of convergence envisions will result in convergence-justified laws. This becomes clear as one considers the pressure constituents bring to bear on their representatives. My criticism of K&S has implications beyond their work: many political liberals, both those who advocate convergence justification and those who do not, draw a similarly problematic distinction between the norms that apply to citizens and those that apply to public officials (Bardon 2015; Habermas 2006; North 2012; Vallier 2014, 181–96, 2015).

For K&S, citizens "do not legislate," and the obligation to enact convergence-justified laws applies to public officials only. K&S do not specify whether convergence justification implies any particular norms for citizens' political activities, since their model ignores them (citizens only "obey"). But other advocates of convergence are clearer: it is a positive implication of

[2] This does not require a romantic conception of democracy where universal activism affects representatives' choices: it is valid even if the main influences come from wealthy, well-connected constituents (Barber 2016).

[3] There are also situations in which cheap talk may contribute to stable social cooperation in iterated games (Skyrms 2004, 65–81).

https://doi.org/10.1017/S000305541700065X Published online by Cambridge University Press

the convergence conception that it does not impose a "restraint requirement" on normal citizens' political advocacy. This allows convergence to be attractive to religious citizens and others who may feel that a requirement that citizens only support properly justified laws conflicts with their "integrity." Because convergence holds that a law is justified when all citizens have a conclusive reason to endorse it, citizens' nonpublic reasons are relevant to convergence justification, as is their comprehensive discourse and advocacy (Gaus 2010; Vallier 2012, 2014, 2015; Gaus and Vallier 2009).

Not all laws are convergence justified; there are many laws that some citizens lack conclusive reasons to endorse. Clear examples include paternalistic policies that interfere with citizens' crucial life choices. Laws forbidding gay marriage, for example, likely cannot be convergence justified, since LGBTQ citizens are part of the public to whom such laws must be justified, and they lack any conclusive reason to endorse them (Gaus and Vallier 2009, 62). Convergence's difficulty arises in explaining how political decision-making processes produce convergence-justified laws when the obligation to only support convergence-justified laws does not constrain ordinary citizens.

Consider K&S's Christian representative. Suppose that a strong majority of her constituents is committed to politically enforcing conservative sexual norms. They support laws that forbid same-sex marriages. K&S's Christian representative, however, knows that such laws are not convergence justified and that she cannot, consistent with her obligations as a public official, support them. So she does not. This angers her constituents, and they withhold support from her, organize campaigns opposing her in the next election, and so on. Because of the hard distinction between the norms that apply to citizens and to public officials in convergence theories, such situations will not only be quite common, empirically, but they are also fully consistent with the normative demands of the theory. Convergence requires public officials to stymie their constituents' political demands. If they do so, then convergence-inspired decision-making processes will result in convergence-justified laws, but at the cost of any meaningful connection between enacted law and citizens' political will. What is more likely, however, is that citizens will eventually find representatives who will accede to their demands: elections do enable citizens to shape laws in the long term (Burstein 2003). And then the laws that representatives produce will not be convergence justified. Without such laws, the state will not treat citizens justly (given convergence's understanding of just), undermining the conditions for assurance and stability. Thus convergence has a democracy problem. This criticism applies by implication to most theories that exempt ordinary citizens from public justification norms, however public justification is defined (Bardon 2015; Habermas 2006; North 2012; Vallier 2014, 181–96, 2015).[4] The problem arises not only from the convergence theory but also whenever theorists draw too hard a distinction between the norms that apply to citizens and to public officials.

Some advocates of convergence think that certain political institutions create situations in which citizens' unjustified inputs produces justified outputs, as if public justification could be analogized to theories of market efficiency. The idea is that American-style, bicameral presidential systems create so many choke points that, by the time legislation is passed, every citizen must have a conclusive reason to endorse it (Gaus and Vallier 2009, 66–71). But this is idealistic in the extreme: it relies on a superficial analogy from economics to politics and ignores the possibility of institutional failure, particularly when convergence explicitly allows advocacy for unjustified laws. Other advocates of convergence may be tempted to claim that they meant their analysis to occur within the confines of ideal theory. Therefore, the criticism that convergence does not offer a plausible explanation of how convergence-justified laws result from citizens' activism ignores their presupposition that citizens and public officials all have active senses of justice. But this is to include an obligation that citizens only support publicly justified laws into convergence's account without arguing for it: my criticism imagines a world in which everyone follows their moral and political obligations as convergence understands them. The problem is that according to convergence theories, citizens are not obligated to support only convergence-justified laws and that advocates of convergence have ignored citizens' influence on the democratic decision-making process.

Convergence thus has a democracy problem: when analyzed in light of the assurance necessary for political liberal stability, the theory fails both to explain how convergence discourse assures public officials and how the political decision-making process convergence allows produces in convergence-justified laws that assure citizens of just treatment by their state. Contra K&S's claims, the stability question demands revision of convergence rather than providing a new argument for it.

I am nevertheless grateful to K&S for contributing to the discussion of political liberal stability. The question of how citizens ought to create enduring relationships of civic trust in the face of deep moral and religious pluralism remains crucial today. The analysis above shows that if political theorists wish to develop persuasive answers to that question, they need to get their hands dirty in careful analysis and understanding of how pluralism, civic society, and democratic political institutions interact (Levy 2015; Waldron 2016). They cannot ignore the democratic context in which these questions are inevitably hashed out, as K&S and other advocates of convergence do, for this will not provide insight on today's crucial political questions. It is important to be reminded of that fact.

[4] Unlike the others above, Habermas offers a theory of how unjustified inputs are deliberatively transformed into justified outputs. I have developed an alternative (2015).

REFERENCES

Barber, Michael Jay. 2016. "Representing the Preferences of Donors, Partisans, and Voters in the U.S. Senate." *Public Opinion Quarterly* 80: 225–49.

https://doi.org/10.1017/S000305541700065X Published online by Cambridge University Press

Bardon, Aurélia. 2015. "Religious Arguments and Public Justification." In *Religion, Secularism, & Constitutional Democracy*, eds. Jean L. Cohen and Cecile Laborde. New York: Columbia University Press, 273–92.
Binmore, Ken. 2005. *Natural Justice*. New York: Oxford University Press.
Burstein, Paul. 2003. "The Impact of Public Opinion on Public Policy: A Review and Agenda." *Political Research Quarterly* 56 (1): 29–40.
Eberle, Christopher J. 2011. "Consensus, Convergence, and Religiously Justified Coercion." *Public Affairs Quarterly* 25 (4): 281–303.
Gaus, Gerald. 2010. "The Place of Religious Belief in Public Reason Liberalism." In *Multiculturalism and Moral Conflict*, eds. Maria Dimova-Cookson and Peter Stirk. New York: Routledge, 19–37.
Gaus, Gerald. 2011a. *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. New York: Cambridge University Press.
Gaus, Gerald. 2011b. "A Tale of Two Sets: Public Reason in Equilibrium." *Public Affairs Quarterly* 25 (4): 305–25.
Gaus, Gerald, and Kevin Vallier. 2009. "The Roles of Religious Conviction in a Publicly Justified Polity: The Implications of Convergence, Asymmetry, and Political Institutions." *Philosophy and Social Criticism* 35 (1-2): 51–76.
Habermas, Jürgen. 2006. "Religion in the Public Sphere." *European Journal of Philosophy* 14 (1): 1–25.
Hadfield, Gillian K., and Stephen Macedo. 2012. "Rational Reasonablenss: Toward a Positive Theory of Public Reason." *Journal of Law and Ethics of Human Rights* 6 (1): 1–46.
Hertzberg, Benjamin R. 2015. "Chains of Persuasion in the Deliberative System: Addressing the Pragmatics of Religious Inclusion." *Journal of Politics* 77 (4): 889–900.
Kogelmann, Brian, and Stephen G. W. Stich. 2016. "When Public Reason Fails Us: Convergence Discourse as Blood Oath." *American Political Science Review* 110 (4): 717–30.
Levy, Jacob T. 2015. *Rationalism, Pluralism, and Freedom*. New York: Oxford University Press.
Lister, Andrew. 2011. "Public Justification of What? Coercion vs. Decision as Competing Frames for the Basic Principle of Justificatory Liberalism." *Public Affairs Quarterly* 25 (4): 349–67.
North, Richard. 2012. "Public Reason, Religious Restraint and Respect." *Philosophia* 40 (2): 179–93.
Rawls, John. 1999. *A Theory of Justice*. Revised ed. Cambridge, MA: Harvard University Press.
Rawls, John. 2001. *Justice as Fairness: A Restatement*. Cambridge, MA: Harvard University Press.
Rawls, John. 2005. *Political Liberalism*. Expanded ed. New York: Columbia University Press.
Skyrms, Brian. 2004. *The Stag Hunt and the Evolution of Social Cooperation*. New York: Cambridge University Press.
Thrasher, John, and Kevin, Vallier. 2015. "The Fragility of Consensus: Public Reason, Diversity and Stability." *European Journal of Philosophy* 23 (4): 933–54.
Vallier, Kevin. 2011. "Convergence and Consensus in Public Reason." *Public Affairs Quarterly* 25 (4): 261–79.
Vallier, Kevin. 2012. "Liberalism, Religion, and Integrity." *Australasian Journal of Philosophy* 90 (1): 149–65.
Vallier, Kevin. 2014. *Liberal Politics and Public Faith: Beyond Separation*. New York: Routledge.
Vallier, Kevin. 2015. "Public Justification versus Public Deliberation: The Case for Divorce." *Canadian Journal of Philosophy* 45 (2): 139–58.
Waldron, Jeremy. 2016. *Political Political Theory: Essays on Institutions*. Cambridge, MA: Harvard University Press.
Watson, Lori. 2015. "Review of Kevin Vallier, *Liberal Politics and Public Faith: Beyond Separation*." Notre Dame Philosophical Reviews. September 20. https://ndpr.nd.edu/news/60955-liberal-politics-and-public-faith-beyond-separation/
Weithman, Paul. 2010. *Why Political Liberalism? On John Rawls's Political Turn*. New York: Oxford University Press.
Weithman, Paul. 2011. "Convergence and Political Autonomy." *Public Affairs Quarterly* 25 (4): 327–48.

https://doi.org/10.1017/S000305541700065X Published online by Cambridge University Press