Do We Know How to Implement Rawls's Liberal Principles of Justice?

HYUNSEOP KIM

Abstract

In this paper, I argue that we do not know how to implement abstract principles of liberal egalitarian justice. Starting with Scheffler's Rawlsian diagnosis of the retreat of liberal democracy in the United States, I argue that it may be due to our lack of knowledge about how to institutionalize a Rawlsian just liberal society. To illustrate the difficulty or challenge, I examine several policy proposals to help build human capital for property-owning democracy and argue that they can fail for various reasons. The main problem is that the changing ways in which diverse individuals respond to policies and interact with one another affect policy consequences, but their complexity surpasses our limited knowledge. The ignorance gives us reason to be patient with the slow pace of building an ideal liberal society, tolerant of those who are sceptical about interventions to implement liberal egalitarian principles, and open to policy experimentation and learning. I further argue that we should publicly acknowledge our ignorance about policy outcomes, as it can reduce political polarization, by moderating policy positions and interpreting policy disagreements as empirical rather than moral, and counter democratic backsliding.

1. Scheffler's Rawlsian Diagnosis of the Democratic Backsliding in the US

Liberal democracy was on the rise in the late 20th century. Now the trend seems to have reversed. In the 21st century, many democracies around the world are backsliding. According to the 2023 report of the Varieties of Democracy (V-Dem) Institute, due to the global decline of democracy, the world now has more closed autocracies than liberal democracies – for the first time since 1995. While the number of democratizing countries is decreasing, the number of autocratizing countries is increasing, and this trend is continuing and worsening (Varieties of Democracy Institute, 2023, pp. 9–14, 19–21; also see Freedom House, 2023; Economist Intelligence Unit, 2022).

doi:10.1017/S0031819125100958 © The Author(s), 2025. Published by Cambridge University Press on behalf of The Royal Institute of Philosophy. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

In 'The Rawlsian Diagnosis of Donald Trump', Scheffler argued that this departure from the ideal of liberal democracy of many liberal democratic countries poses a challenge to liberalism as a political philosophy (Scheffler, 2019). When sustained attempts to realize a political theory in practice run into serious trouble, we should pause to wonder whether the political theory is feasible and deserves our effort for its realization. To allay our concern over the feasibility of liberalism, liberal theorists owe an explanation of why actual societies, which have, at least apparently, attempted to realize their liberal theory, have failed to conform to the requirements of their theory.

In the case of the US, Scheffler's Rawlsian diagnosis is that its institutions and policies have egregiously failed to realize the ideal of reciprocity in recent decades. Reciprocity is what makes Rawls's architecture of a liberal society just and stable. As people cooperate with one another on mutually advantageous terms within a just basic structure, they develop allegiance to liberalism and its institutions and trust and confidence in their fellow citizens over time. In particular, when the advantaged support the difference principle and avoid earning more at the expense of the disadvantaged, the public affirmation of concern and solidarity fosters civic friendship and stabilizes the considerate rules of social cooperation. However, the US in recent decades has fallen far too short of this ideal. Most of all, as the rising economic inequality attests, Scheffler argued, its economic institutions have enabled the wealthy and privileged to maximize their lot at the expense of everyone else. The US institutions have blatantly ignored, shamelessly disregarded and failed to even strive seriously to achieve the ideal of reciprocity and have instead embraced staggering levels of inequality. So, it is not surprising that the grievance, anger, and resentment of the unfairly disadvantaged at the dysfunctional social system, exploited by demagogues, wound up electing a president who led the country even further away from the ideal of liberal democracy (Scheffler was referring to the first election of Donald Trump in 2016). As it was not the theory's defect but the US society's failure to implement it properly that caused the trouble, Rawls's theory is off the hook. Indeed, Rawls's theory is vindicated as it provides theoretical resources to diagnose the malaise of American society and explain what went wrong.

I agree with Scheffler that Rawls's moral-political psychology in his account of the ideal liberal society's stability offers a model of harmonious social relationship by reference or contrast to which we

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

can explain the social psychological mechanism of democratic backsliding. Indeed, the diagnosis seems applicable not only to the US but to many other parts of the globe. Scheffler's indictment of the US's blatantly unfair socio-economic system seems to describe the perception of many people of their own socio-economic system. Their country, the perception goes, shamelessly disregards many ordinary citizens' interests, blatantly ignores the call for fairness, and does not even try to achieve the goal of equality and justice. Their thought, as the strong words of condemnation suggest, may be that there exist sets of just institutions and policies near at hand that, once adopted, will realize liberal principles of justice. It is only a matter of political will, or of overcoming the resistance of those handful of people who, presumably for their own benefit, stand in the way of implementing the just institutions and policies.

While their perception of their society is understandable, I am afraid that the thought, especially that there are easily available ways of implementing liberal principles of justice, may not be entirely correct. Our knowledge, at the institutional and policy level, of how to realize a just liberal society is rather limited. The problem is more epistemic than volitional. Institutions and policies often do not yield the outcome that their makers intend to achieve, not least because those who are subject to the institutions and policies respond to their incentives in hardly predictable ways. Even if all of us were to agree to liberal principles of justice, say, Rawls's principles of justice as fairness, we should expect there to be reasonable disagreement as to how to institutionalize them. Failing to appreciate the extent of our ignorance in policymaking leads us to hold extreme policy positions and contributes to policy polarization. Moreover, the ignorance of our ignorance easily makes our discontent at the existing order of things, which itself may be reasonable, impatient and intolerant. Impatient resentment and indignation are liable to be channelled in unpredictable directions and exploited, by demagogues, in politically dangerous ways. Populist politicians can transform people's amorphous discontent at the *status quo* into animosity against their opposite party and generate affective polarization. When citizens are highly polarized, they can prioritize their partisan interests over democratic principles and vote for undemocratic candidates who tilt the political playing field in their favour. So, it is important to publicly acknowledge our limited knowledge about institutions and policies, because it would help to slow, if not stop or reverse, democratic backsliding. Or so I will argue in the remainder of this paper.

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

2. Do We Know How to Implement Rawls's Property-Owning Democracy?

Suppose that all citizens in your society sign on to Rawls's two principles of justice as fairness and deliberate about how to design institutions and policies to realize the principles.[1] While many people (mis)interpret Rawls's principles of justice as a defence of welfare state institutions and policies, Rawls himself argues that his principles of justice can hardly be realized by a welfare state that only aims to assist those who lose out through accident or misfortune and make sure that no one falls below a decent minimum standard of living. When generous welfare provisions generate 'a discouraged and depressed underclass many of whose members are chronically dependent on welfare', the redistribution of income is hardly an expression of the principle of reciprocity. Reciprocity can be realized in the regime of property-owning democracy (for short, POD) that ensures the widespread ownership of productive assets and human capital (= educated abilities and trained skills). When most people command sufficient amounts of productive means, they can take part in mutually advantageous social cooperation on a footing of suitable equality and mutual respect (Rawls, 1999, pp. xiv–xv; Rawls, 2001, pp. 137–40). They are also secure, independent, and free from domination, since they are not subject to the whim of an employer for an income or livelihood (Meade, 1964, p. 39).

So, the citizens in your society decide that, to implement POD and achieve widespread ownership of productive capital, the government help under-capitalized people build human capital.[2] There are several reasons why they focus on human capital (rather than

[1] In a liberal society, we should expect disagreement about principles of justice. Rawls admits that it would be unreasonable for him to deny that there are other reasonable liberal political conceptions of justice that do not include the difference principle but replace it with a less egalitarian principle (*e.g.*, the principle of maximizing average utility, on condition of guaranteeing for everyone an adequate level of minimum income) (Rawls, 1993/2005, p. xvlii). Since this paper addresses the disagreement about translating a liberal conception of justice into institutional designs and policy proposals, the reasonable disagreement about which liberal conception of justice is the correct one is set aside.

[2] Rawls leaves open the possibility that liberal socialism realizes his principles of justice. However, the knowledge problem of liberal socialism would presumably be worse than the knowledge problem of property-owning democracy that I will explain, because the former would need greater centralization of state power and bureaucracy for operation (if it

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

physical capital). First, human capital enables the possessors to work effectively, acquire a sense of efficacy, and gain recognition from others. Self-efficacy and recognition are an important source of self-respect, a major component of Rawls's primary goods. When they exercise their abilities at work and thereby receive recognition, the work is meaningful and gives them a sense of fulfilment (Veltman, 2016, Ch. 5).[3] Second, individuals may not take into account the social values of human capital (*e.g.*, knowledgeable and skilled workers increase the productivity of their coworkers by helping the firm they work for adopt productive technologies and create new technologies) and underinvest in their human capital. Also problematic is the imperfect credit market for human capital. The return on human capital investment is highly uncertain, partly because human capital, since embodied, can hardly be offered as collateral for a loan (Becker, 1964/1993, pp. 91–4). The credit constraint can be particularly severe for the poor. So, there is room for policy interventions to correct the underinvestment in human capital (especially of disadvantaged children), make the capital distribution less uneven, and foster economic growth (Galor and Zeira, 1993; Aghion and Bolton, 1997; Galor and Moav, 2004). Third, the skill-biased technological change in recent decades has made human capital more important and indispensable for full participation in the knowledge economy (Machin and Van Reenen, 1998; Acemoglu, 2002).

However, the challenge is how to build solid institutions and implement effective policies that actually work as intended. Suppose the citizens in your society exercise their political will and collect a large revenue by taxing the rich and privileged. Imagine you are the prime minister who is in charge of helping the under-capitalized build human capital. How would you spend the money to bring about the desirable state of affairs in which most people possess sufficient amounts of human capital?

---

operates at all) than the latter. Thomas argues that market socialism, without an efficiently functioning labour market, would undermine the freedom of occupation and violate Rawls's principle of reciprocity (Thomas, 2016, Ch. 8).

[3] Using knowledge and skills at work also enables the worker to acquire related knowledge and skills and take on more complex roles. It can create the virtuous circle of capacity building and role enhancement, which makes work enjoyable and energizing. *Cf.* Rawls's Aristotelian Principle (1999, pp. 373–4).

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

Subsidizing education or schools may suggest itself. However, more resources in schools do not necessarily mean better education. Building school facilities, reducing the number of students per class or the teacher-student ratio, or hiring teachers with advanced degrees may not much improve the quality of student learning (Hanushek, 2003). Just paying teachers handsomely might not solve the problem, either. It is hardly predictable how the teachers will respond to monetary incentives. Will they be incentivized to improve their pedagogy and take better care of students? Might it not instead attract those who care more about financial rewards than students?

How about subsidizing college tuition, especially for the students from low-income families? It may increase college enrolments, but the increased enrolment does not necessarily mean increased human capital. According to the signalling theory of education, college graduates earn higher incomes, not because they acquire productive skills and knowledge at the colleges, but rather because a college degree works as a signal of the productivity that its holder had prior to going to college. The diploma indicates to potential employers that its holder is likely to be more intelligent, diligent, and disciplined than other candidates who do not have one. If schooling is valuable as a signal, it would be a positional good. As more people have the signalling good, its function of distinguishing its possessor from others diminishes. When having a college diploma becomes the norm, not an exception, it cannot put its holder ahead of others in the labour market. If everybody has one, it does not signal anything. The arms race does not make any country safer or stronger. As the value of armaments lies mainly in giving the armed country a competitive edge over others, the relative advantages cancel each other out. Similarly, if and to the extent that the value of college attendance lies in the certificate that gives its possessor relative advantages, the money spent on subsidizing college tuition is wasted in vain.[4]

One might argue that you can and should spend more wisely by subsidizing education not across the board, but only of those disciplines that transmit productive skills and knowledge to students. The problem is that the government can hardly know enough to determine whether and how to restrict subsidies to certain disciplines and exclude others. It is hard to predict which skills and

[4] For the positional aspect of education, see Brighouse and Swift (2006) and Caplan (2018).

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

knowledge will come in handy during the time when the students will be in the job market. It seems likely that various skills and knowledge from many disciplines would combine together and contribute in unpredictable ways. (Who could have predicted that Steve Jobs's study of calligraphy in college would pay off?) Neither does raw data itself tell us which disciplines are more productive and worthy of more investment than others.[5]

Arguably, a better way of implementing POD is to decentralize human capital investment decisions. Decentralization enables individuals to make use of their local knowledge, what they know about themselves and their specific circumstances. School voucher programs funded by tax revenues can be a policy that makes use of individuals' knowledge of their own potential, needs, and wants. The idea is that empowering students to choose schools enables them to attend the schools that fit them and encourages schools that meet their diverse needs and wants to emerge. Competition among schools pushes schools to respond to students' preferences. However, it might not work as expected. School vouchers can lead instead to the sorting of students across parental income and student talent. Such sorting may adversely affect disadvantaged students due to peer effects and aggravate social segregation.[6]

A radical policy proposal in this vein is to give a generous amount of cash to young adults that they can use as they see fit. The intent or hope is that they invest the basic capital wisely, say, in starting their own businesses, enrolling in apprentice training programs, or whatnot. Society should encourage young minds to form an ambitious plan of life without getting distracted by an urgent need to make ends meet. They should be given an adequate opportunity to sincerely ask 'what do I really want out of my life?' at the start of their adult life (Ackerman and Alstott, 1999; White, 2003, Ch. 8). The basic capital policy respects young adults' autonomous choice of a life project. It can help them develop and exercise what Rawls calls the capacity for a conception of the good (Rawls 2001, p. 19). Unlike subsidizing higher education, it does not disadvantage those who are not academically inclined and thereby aggravate socioeconomic

[5] American Philosophical Association points out that philosophy majors earn more than other humanity majors: https://www.apaonline.org/page/data. Is it because philosophy education in colleges develops productive skills and knowledge, or because those who already have productive skills and knowledge choose to major in philosophy?

[6] For the arguments for and against school vouchers and mixed empirical results, see Epple *et al*. (2017).

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

inequality (White, 2010). Despite these virtues, an immediate worry is that young adults may not be mature enough to handle a large sum of money. They can squander it by purchasing luxuries or investing in a failing business. On the other hand, if they turn out not to be adventurous and choose the safe and seemingly reasonable bet, *i.e.*, using the money to pay for college tuition, the problem of positionality in education arises (again). A large portion of what makes college attendance good for a student might be its signalling value. Whether college education builds human capital or merely sends a signal does not make a difference to whether the investment pays off for the individual student. However, if many young adults act in their own interest and engage in signalling that counteracts each other, their basic capital is wasted in sending mutually cancelling signals.

I have introduced and critically examined several possible ways of implementing POD. My goal is not to weigh in on, let alone settle, the debate about whether and how we can implement POD. I have taken this example to illustrate the difficulties in designing or finding suitable institutions and policies that realize abstract principles of social justice. I readily admit that the examination is neither complete nor exhaustive. While these policies can fail for the reasons examined above, they might nonetheless work in some places, at some times. There might be other kinds of institutions and policies and combinations thereof that can get around the challenges. The problem is that we often do not have sufficient evidence to tell, among many possible candidates, which policy would work here and now. The reason why we, in general, do not know the consequences of policies, which applies to many policies for POD above, can be stated as follows.

The socioeconomic facts that affect policy consequences are numerous, diverse, and complicated. The outcome of policy implementation depends not only on the ever-changing environment (including technological innovations), but also on the ways in which many individuals respond to the policy and interact with one another. Achieving the goal of legislation often requires coordination of many people's behaviour, but the government cannot expect people to behave as it wants them to. It is ignorant and arrogant for policymakers to imagine that individuals are simple creatures whose actions they can easily predict and, with their policy tools, lead in the directions they want (*cf.* Smith, 1759/2002, p. 275). In a liberal society, different people have different abilities, values, and plans of life, living under different circumstances. In many strategic situations, one's action depends on her expectation of others' actions, which in

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

turn depend on their expectations of her action and the actions of still others, including policymakers. Inevitably inaccurate expectations and multiple Nash equilibria of people's preferences, strategies, and their interactions with one another and with institutions make it quite challenging to predict how people will react to a policy and how the policy will pan out.

However, we usually do not have sufficient information. Our social scientific understanding is incomplete in data and theory. Policy outcomes depend not only on general facts that social sciences can in principle capture with abstract models and statistical analysis on aggregate data, but also on particular facts about local circumstances. Information about particular facts is fragmented and dispersed among many individuals. What policymakers know about individuals' particular circumstances is limited (Hayek, 1945). Indeed, the government's knowledge should be limited. The government should respect citizens' right to privacy. The government should not resort to coercive or manipulative measures to extract private information. If a policy has been tried in other societies, it can help to examine how the policy has worked in those societies. However, an intrinsically identical policy might yield different results, depending on whether other institutions and people's dispositions are complementary or incompatible. The factors that make the policy effective in one society may be absent in another society, and the policy can cause unexpected, undesirable side effects in the other society that more than offset its beneficial effects (Cartwright and Hardie, 2012).

So, it is not surprising that policy outcomes are subject to reasonable disagreement. By reasonable disagreement, I mean the disagreement about what the society's institutions and policies should be among the decent, well-intentioned people who reason about policy outcomes in good faith and to the best of their abilities. They are well-meaning people who are willing to cooperate with their fellow citizens on fair, mutually beneficial terms. At the same time, as we all do as epistemically limited creatures, they interpret the world through the lens of their experiences and use epistemic heuristics or shortcuts. Due to differences in past experiences, they use different schemas and theories that yield different predictions about policy outcomes. As 'the evidence – empirical and scientific – bearing on the case is conflicting and complex, and thus hard to assess and evaluate,'[7] it is hard to tell which of them, if any, is (largely) correct.

[7] It is one of what Rawls calls 'the burdens of judgment' that generate reasonable disagreement (Rawls, 1993/2005, p. 56). While Rawls

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

As they make different predictions about what the consequences of policies would be, they can end up supporting different policies.

3. Practical Lessons: Patience, Tolerance, and Openness to Policy Experimentation

I have argued that we hardly know how to implement Rawls's regime of property-owning democracy. Even if all of us were to agree to Rawls's principles of justice, we should expect there to be reasonable disagreement as to how to institutionalize them. What practical implications does our ignorance about policy outcomes, in particular about how to realize liberal principles of justice, have? What practical lessons can we draw?

The government's knowledge problem significantly limits its capacity to bring about what it intends to achieve. So, we should not be surprised or overly frustrated when it takes time and trial and error for the government to achieve a demanding goal (such as realizing the principles of liberal egalitarianism), especially when the goal involves transformation of citizens themselves (as is the case with the establishment of a POD). It is a daunting task. Do not get me wrong. I do not mean that we should give up the project of implementing POD or building a liberal egalitarian society. I believe that we should not be discouraged by the possibility and experience of failure. As we do not know whether the project will succeed, neither do we know whether it will fail. In my view, we have not tried long and hard enough to conclude that POD or a liberal egalitarian society is an unattainable utopia. At the same time, we have reason to be patient.

We also have reason to be tolerant. When we attempt to build a regime or a just society that has never been actually built and run, we may not know how we can build one. Even those of us who support the regime might not know for sure whether we can or whether it is feasible at all. So, it should not come as a surprise if other people are sceptical. Their scepticism might be reasonable and, given their life experiences and worldview, may well be at least understandable. As I have argued, people can easily and reasonably disagree about whether some policy proposals will work and what

---

deals mainly with the axiological disagreement about values in life in *Political Liberalism*, here I focus on the empirical disagreement about policy outcomes.

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

they will achieve. The interactions between policies and individuals' responses are enormously complex and all of us are subject to ignorance about such complicated matters. It may be the case that those on the other political side share the same or at least similar values and ideals – liberal democracy and some form of egalitarian liberalism – with us. The disagreement may be about what those values and ideals require us to do and politically support, *i.e.*, their application, not about fundamental values. We may not know whether the disagreement is due to a conflict about fundamental values, or to different calculations about how to realize shared values. In this situation, it may be premature to conclude that those whose policy preferences are different from ours do not care about justice and equality. They might be supporting other ways of realizing liberal principles of justice (*e.g.*, Tomasi, 2012, Ch. 8). It may be groundless to accuse them of greed or a penchant for domination without further evidence. It would be unreasonable to scorn, vilify, or demonize them. Toleration would be in order.

Furthermore, we have reason to work together to reduce our empirical ignorance. Fortunately, disagreement due to empirical ignorance is less intractable than disagreement about fundamental values. We can conduct policy experiments that both advocates and sceptics can agree on and see where they lead us. When a policy that has not been tried and tested is put on the table, many people can be sceptical whether it would work as intended. Their scepticism may be reasonable or at least understandable. Still, when the sceptics realize that as we do not know whether the policy will succeed, neither do we know whether it will fail, they can be persuaded into taking a small experimental step and seeing where it leads. Indeed, it may be unreasonable to refuse a policy experiment when the proposal makes sense in theory and seems worthy of experimentation. Advocates of the policy also can settle for a small-scale trial, when there is scant evidence to justify a big policy commitment upfront. It may be wise and more effective in the long run to begin with a pilot implementation, improve the policy through trial and error, gain support by demonstrating that it works (if it does), and scale up with more support. Indeed, it is arguably illegitimate to impose an untested, large-scale policy on others when they are reasonably sceptical (*cf.* Rawls, 1993/2005, p. 137; Gaus, 2011, pp. 490*ff.*).

So, both advocates and sceptics can and should agree to experiment with policies, first on a small scale, and learn from failure or success. Either way, the evidence from experimentation can make all of them less ignorant and help make effective policies. Experimenting with a policy in one's society is more valuable than

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

observing how the policy worked in other societies, as the natural environment, institutions, other policies, and people's dispositions are controlled for in policy experimentation in one's own society. As people learn from experiments and adjust policies accordingly, they can acquire knowledge and develop relevant virtues. They can learn social sciences (*e.g.*, institutional economics, statistics) and overcome the temptation to (mis)interpret results to rationalize their *ex-ante* policy position. A virtuous circle can develop: policy experiments, improvement in policy outcomes, increase in people's knowledge and virtues, policy experiments on a bigger scale and in more areas. The learning experience can help people understand each other, reduce their policy distance, and build mutual trust. As this beneficial effect has to do with democratic backsliding, it is worthy of elaboration in the next section.

4. How the Public Understanding of our Ignorance About Policy Outcomes Can Alleviate Political Polarization and Counter Democratic Backsliding

Research in cognitive science shows that we tend to overestimate our understanding of how everyday objects like toilets, bicycles and combination locks work. When we try to explain their causal mechanism, we realize their complexity and recognize that our sense of understanding was an illusion (Rozenblit and Keil, 2002). Similarly, we often mistakenly believe that we know the consequences of implementing policies. Fernbach and Sloman demonstrated the illusion of understanding policy outcomes through the following experiment: They first asked participants to rate their understanding of various policies. Next, they asked them to provide an explanation of all the consequences that one of the policies would lead to. Participants were then asked to re-rate their understanding of the policy. It turned out that they rated their understanding lower than in the initial assessment. They also asked the participants to rate their position (support or opposition) on the policy on a scale of 1–7. Participants stated their position, before and after they tried to explain what the consequences of a policy would be. Attempting to explain the policy's mechanism reduced not only their sense of understanding of the policy, but also the certainty and extremity of their policy position. The moderating effect also manifested in political behaviour. After rating their position on a given policy and providing a causal explanation of it, some participants chose whether to donate a bonus payment to a relevant advocacy group.

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

Those who initially held a strong position donated no more often than those who were initially more moderate (Fernbach *et al*., 2013).

You might think that I just gave you a practical reason *not* to emphasize the difficulty of predicting policy outcomes. We need more donations to advocacy groups, not less. Being moderate is neither fun nor cool. It will probably not make you popular in this polarized era. Let me introduce another experiment in political psychology and explain why we have reason to be moderate. Democratic backsliding is a puzzling phenomenon. It is a process that starts from a democratic *status quo*. In a democracy, people have the ultimate political power; institutionally, they have the right to vote. And most people support democracy. Then, how can it happen that they let democracy backslide? Why don't they vote against those politicians who undermine democracy? If the incumbent subverts democracy, why don't ordinary people vote him out of power? A plausible hypothesis is that in a deeply polarized society, many voters who value democracy put their partisan interests, which include policy preferences, before democratic principles. They do not withdraw support from the incumbent who tilts the political playing field in his favour, but they tolerate him as he champions the policies they prefer. They cannot bring themselves to vote for the other candidate whose policies they strongly oppose, even when they know that the incumbent undermines democracy (Svolik, 2019).

Svolik's survey experiment corroborated this hypothesis. A representative sample of American voters were asked to choose between hypothetical candidates, some of whom endorsed undemocratic positions (*e.g*., gerrymandering, voter suppression, the executive's prosecution of critical journalists, and ignoring court rulings). The share of votes undemocratic candidates received was compared with the share of votes democratic but otherwise identical candidates received. The decline in undemocratic candidates' vote share can be a metric for the punishment voters would dispense at the ballot box in defence of democracy. The result was not encouraging for democracy. It seemed that many Americans would vote for candidates who represent the socioeconomic policies they prefer, even if those candidates violate democratic principles. Voters punished violations of democratic principles less severely than extramarital affairs. When candidates from their favourite party violated undemocratic principles, it had a smaller impact than when candidates from the other party exhibited such behaviour. Most importantly to my point, the larger the difference between the voter's policy preferences and the candidate's, the weaker the tendency to prioritize democratic principles in their electoral choices. So, centrist voters who see

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

small policy differences between candidates punish undemocratic behaviour much more severely than extremist voters who strongly favour one of the candidates (Graham and Svolik, 2020). To sum up: when we realize our ignorance about policy outcomes, our policy positions become moderate. The moderation reduces *policy* polarization, the divergence of policy preferences among voters. It can help to prevent policy polarization from trumping democratic principles and counter democratic backsliding.

Moreover, we have practical reasons to widely publicize and publicly recognize that we are all largely ignorant about policy outcomes. When we know our own ignorance about policy outcomes, know further that others also know their own ignorance, and openly recognize the knowledge of ignorance (and the knowledge thereof), the commonly known, shared ignorance can bring about good consequences. (1) We tend to overestimate the extremity of others' policy positions, and the overestimation can make our own policy position more extreme (Ahler, 2014; Levendusky and Malhotra, 2015). So, if we publicly acknowledge our common ignorance about policy outcomes and see others moderate their policy position, it can correct our misperception of policy polarization and reduce actual policy polarization. (2) It can also reduce *affective* polarization, in particular the anger and hatred toward the opposing party and its supporters. When we see those who initially disagree with us move their policy position toward ours, perceived policy distance shrinks. As perceived policy distance contributes to affective polarization (Webster and Abramowitz, 2017; Dias and Lelkes, 2022), affective polarization decreases accordingly. Moreover, partisans often misbelieve that their adversaries work to undermine the values they care deeply about, and the misapprehension correlates with disliking of the outgroup (Chambers and Melnyk, 2006). If deliberation over policy outcomes reveals to us that much of the disagreement over policies has to do with empirical uncertainty that all of us are subject to because of the causal complexity of polices and our limited knowledge, we would perceive less disagreement over values and ideals. For example, liberals might realize that many conservatives favour a less progressive income tax, not because they do not care about the poor, but because they believe, rightly or wrongly, that reducing progressivity will increase entrepreneurship, boost economic growth, and eventually make everyone better off. Similarly, the reason why conservatives support school vouchers can turn out to be not because they want to undermine social integration, but because they believe, rightly or wrongly, that it will make schools better and improve students' academic performance. It can help

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

reduce partisan animosity if we see that those on the other political side are not opposed to the core values that we cherish. It can help restore civic friendship when policy disagreements are attributed not to fundamental value conflicts, but to different calculations about policy outcomes. Reduction of affective polarization, as it improves the relationship among fellow citizens, is in itself good. It might also help counter democratic backsliding (Kingzette *et al*., 2021). (3) As I have argued in the previous section, public recognition of our common ignorance about policy outcomes can lead us to conduct policy experiments. The results can further reduce political polarization.

5. Conclusion

Let me take stock. I have illustrated, with the example of policies to implement the regime of property-owning democracy, the difficulty of realizing Rawls's liberal principles of justice. The main challenge lies in predicting policy outcomes. The changing ways in which diverse individuals respond to policies and interact with one another affect policy consequences, but their complexity surpasses our limited knowledge. Our ignorance about policy outcomes gives us reason to be patient with the slow pace of building a Rawlsian ideal society, tolerant of those who are sceptical or cautious about interventions to implement liberal egalitarian principles, and open to policy experimentation and learning. We have reason to publicly recognize our ignorance about policy outcomes, as it can counter political polarization and democratic backsliding. Democratic backsliding or the electoral wins of undemocratic politicians such as Donald Trump may be due not only to our failure to realize liberal egalitarianism, but also to our failure to recognize and appreciate the difficulty of its realization and resultant impatience and intolerance.

If my arguments are correct, is it good news for Rawlsian liberalism? Scheffler argues that when attempts to realize a political theory fail over an extended period of time, its supporters owe an explanation about the failure and defend the theory's feasibility or viability. Here is my explanation: Rawls's model of liberal society is ideal in the sense that *once* a society is as well-ordered as the model (*i.e*., '(1) everyone accepts and knows that the others accept the same principles of justice, and (2) the basic social institutions generally satisfy and are generally known to satisfy these principles' (Rawls, 1999, p. 4), the well-ordered society reproduces citizens' compliance with the principles and stably continues to be well-ordered. If, as I have argued, reasonable people can disagree over

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

how to institutionalize Rawls's principles of justice, apparently liberal societies (such as the US) have not actually been well-ordered. The second condition has probably not been met. We have not yet reached a point at which the stability or sustainability of Rawls's ideal society is put to the test. So, the recent departure from the liberal ideal or democratic backsliding does not refute Rawls's ideal theory.

While my analysis suggests that establishing a well-ordered liberal society is a hard-to-achieve goal (maybe, harder than many people have thought), the possibility of achieving or at least getting close to it is not so remote as to make Rawls's ideal theory practically irrelevant. For one, thanks to recent developments in social science and its methods, we have greater knowledge about societies than before (*e.g.*, Abhijit and Duflo, 2019). We can also accumulate knowledge about our own society and incrementally improve policies through trials and errors (Lindblom, 1959). A just liberal society where citizens reciprocally cooperate with one another on fair terms is an attractive ideal that deserves our sustained effort for its realization. The goal of realizing one is still ahead of us, and it is too early to despair. We should, I believe, patiently continue to learn through policy experiments and increase publicly shared knowledge about institutions and policies. Making people realize the complexity of modern liberal societies, our limited knowledge, and our fallibility about policy outcomes can be a good first step to motivate people to engage in this process and facilitate political depolarization.[8]

References

Banerjee Abhijit and Esther Duflo, *Good Economics for Hard Times* (New York: Public Affairs, 2019).

Daron Acemoglu, 'Technical Change, Inequality, and the Labor Market', *Journal of Economic Literature*, 40.1 (2002), 7–72.

[8] Earlier drafts of this paper were presented at the Conference in honour of Sam Scheffler at UC Berkeley and the Eighth Annual Philosophy-Politics-Economics Society Meeting in New Orleans. I would like to thank the audience on those occasions, including Johathan Benson, Daniela Dover, Jed Lewinsohn and Regina Rini, for helpful comments and discussions. This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2022-NR070855).

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

Bruce Ackerman and Anne Alstott, *The Stakeholder Society* (New Haven: Yale University Press, 1999).
Philippe Aghion and Patrick Bolton, 'A Theory of Trickle-Down Growth and Development', *The Review of Economic Studies*, 64.2 (1997), 151–172.
Douglas J. Ahler, 'Self-Fulfilling Misperceptions of Public Polarization', *The Journal of Politics*, 76.3 (2014), 607–620.
Gary Becker, *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, 3rd edn (Chicago: University of Chicago Press, 1993; first publ 1964)
Harry Brighouse and Adam Swift, 'Equality, Priority, and Positional Goods', *Ethics*, 116.3 (2006), 471–497.
Bryan Caplan, *The Case Against Education: Why the Education System Is a Waste of Time and Money* (Princeton: Princeton University Press, 2018).
Nancy Cartwright and Jeremy Hardie, *Evidence-Based Policy: A Practical Guide to Doing It Better* (Oxford: Oxford University Press, 2012).
John R. Chambers and Darya Melnyk, 'Why Do I Hate Thee? Conflict Misperceptions and Intergroup Mistrust', *Personality and Social Psychology Bulletin*, 32.10 (2006), 1295–1311.
Nicholas Dias and Yphtach Lelkes, 'The Nature of Affective Polarization: Disentangling Policy Disagreement from Partisan Identity', *American Journal of Political Science*, 66.3 (2022), 775–790.
Economist Intelligence Unit, *Democracy Index 2022*, available at https://www.eiu.com/n/wp-content/uploads/2023/02/Democracy-Index-2022_FV2.pdf.
Dennis Epple, Richard E. Romano and Miguel Urquiola, 'School Vouchers: A Survey of the Economics Literature', *Journal of Economic Literature*, 55.2 (2017), 441–492.
Philip M. Fernbach et al, 'Political Extremism Is Supported by an Illusion of Understanding', *Psychological Science*, 24.6 (2013), 939–946.
Oded Galor and Omer Moav, 'From Physical to Human Capital Accumulation: Inequality and the Process of Development', *The Review of Economic Studies*, 71.4 (2004), 1001–1026.
Oded Galor and Joseph Zeira, 'Income Distribution and Macroeconomics', *The Review of Economic Studies*, 60.1 (1993), 35–52.
Gerald Gaus, *The Order of Public Reason* (Cambridge: Cambridge University Press, 2011).
Matthew Graham and Milan Svolik, 'Democracy in America? Partisanship, Polarization, and the Robustness of Support for Democracy in the United States', *American Political Science Review*, 114.2 (2020), 392–409.
Stephan Haggard and Robert Kaufman, *Backsliding: Democratic Regress in the Contemporary World* (Cambridge: Cambridge University Press, 2021).
Eric A. Hanushek, 'The Failure of Input-Based Schooling Policies', *The Economic Journal*, 113.485 (2003), F64–F98.

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

F. A. Hayek, 'The Use of Knowledge in Society', *The American Economic Review*, 35.4 (1945), 519–530.
Freedom House, *Freedom in the World 2023*, available at https://freedomhouse.org/report/freedom-world/2023/marking-50-years.
Jon Kingzette et al, 'How Affective Polarization Undermines Support for Democratic Norms', *Public Opinion Quarterly*, 85.2 (2021), 663–677.
Matthew S. Levendusky and Neil Malhotra, '(Mis)Perceptions of Partisan Polarization in the American Public', *Public Opinion Quarterly*, 80.S1 (2016), 378–391.
Charles E. Lindblom, 'The Science of 'Muddling Through'', *Public Administration Review*, 19 (1959), 79–88.
Stephen Machin and John Van Reenen, 'Technology and Changes in Skill Structure: Evidence from Seven OECD Countries', *The Quarterly Journal of Economics*, 113.4 (1998), 1215–1244.
James Meade, *Efficiency, Equality and the Ownership of Property* (London: George Allen & Unwin, 1964).
John Rawls, *A Theory of Justice*, rev. edn (Cambridge, MA: Harvard University Press, 1999).
John Rawls, *Justice as Fairness: A Restatement*, ed. by Erin Kelly (Cambridge, MA: Harvard University Press, 2001).
John Rawls, *Political Liberalism*, expanded edn (New York: Columbia University Press, 2005; first publ. 1993)
Leonid Rozenblit and Frank C. Keil, 'The Misunderstood Limits of Folk Science: An Illusion of Explanatory Depth', *Cognitive Science*, 26 (2002), 521–562.
Samuel Scheffler, 'The Rawlsian Diagnosis of Donald Trump', Boston Review, 12 February 2019, available at https://www.bostonreview.net/articles/samuel-scheffler-rawlsian-diagnosis-donald-trump.
Adam Smith, *The Theory of Moral Sentiments*, ed. by Knud Haakonssen (Cambridge: Cambridge University Press, 2002; first publ. 1759)
Milan Svolik, 'Polarization versus Democracy', *Journal of Democracy*, 30.3 (2019), 20–32.
Alan Thomas, *Republic of Equals: Predistribution and Property-Owning Democracy* (Oxford: Oxford University Press, 2016).
John Tomasi, *Free Market Fairness* (Princeton: Princeton University Press, 2012).
Varieties of Democracy (V-Dem) *Institute, Democracy Report 2023*, available at https://v-dem.net/documents/29/V-dem_democracyreport2023_lowres.pdf.
Andrea Veltman, *Meaningful Work* (Oxford: Oxford University Press, 2016).
Steven W. Webster and Alan I. Abramowitz, 'The Ideological Foundations of Affective Polarization in the U.S. Electorate', *American Politics Research*, 45.4 (2017), 621–647.
Stuart White, *The Civic Minimum: On the Rights and Obligations of Economic Citizenship* (Oxford: Oxford University Press, 2003).

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press

Do We Know How to Implement Liberal Justice?

Stuart White, 'A Modest Proposal? Basic Capital vs. Higher Education Subsidies', *The British Journal of Politics and International Relations*, 12.1 (2010), 37–55.

HYUNSEOP KIM (h.kim@snu.ac.kr) *is Professor of Philosophy at Seoul National University. His recent publications appeared in Political Theory, International Theory, Oxford Studies in Philosophy of Mind, and Economics & Philosophy. He is working on, among other things, property-owning democracy, democratic backsliding, and the value of economic growth.*

https://doi.org/10.1017/S0031819125100958 Published online by Cambridge University Press