

JOURNAL *of* ETHICS
& SOCIAL PHILOSOPHY

VOLUME XVII · NUMBER 2

April 2020

ARTICLES

- 103 Social Reform in a Complex World
Jacob Barrett
- 133 Freedom and Actual Interference
Jonah Goldwater
- 159 Do We Have Reasons to Obey the Law?
Edmund Tweedy Flanigan
- 198 The Ambitions of Consequentialism
Brian McElwee

DISCUSSIONS

- 219 Error Theory, Unbelievability, and the
Normative Objection
Daniele Bruno
- 228 Does Initial Appropriation Create New
Obligations?
Jesse Spafford

The *Journal of Ethics and Social Philosophy* (ISSN 1559-3061) is a peer-reviewed online journal in moral, social, political, and legal philosophy. The journal is founded on the principle of publisher-funded open access. There are no publication fees for authors, and public access to articles is free of charge and is available to all readers under the CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 license. Funding for the journal has been made possible through the generous commitment of the Gould School of Law and the Dornsife College of Letters, Arts, and Sciences at the University of Southern California.

The *Journal of Ethics and Social Philosophy* aspires to be the leading venue for the best new work in the fields that it covers, and it is governed by a correspondingly high editorial standard. The journal welcomes submissions of articles in any of these and related fields of research. The journal is interested in work in the history of ethics that bears directly on topics of contemporary interest, but does not consider articles of purely historical interest. It is the view of the associate editors that the journal's high standard does not preclude publishing work that is critical in nature, provided that it is constructive, well-argued, current, and of sufficiently general interest.

Editor

Mark Schroeder

Associate Editors

Dale Dorsey

James Dreier

Julia Driver

Anca Gheaus

Tristram McPherson

Colleen Murphy

Hille Paakkunainen

Discussion Notes Editor

Kimberley Brownlee

Editorial Board

Elizabeth Anderson	Philip Pettit
David Brink	Gerald Postema
John Broome	Joseph Raz
Joshua Cohen	Henry Richardson
Jonathan Dancy	Thomas M. Scanlon
John Finnis	Tamar Schapiro
John Gardner	David Schmidtz
Leslie Green	Russ Shafer-Landau
Karen Jones	Tommie Shelby
Frances Kamm	Sarah Stroud
Will Kymlicka	Valerie Tiberius
Matthew Liao	Peter Vallentyne
Kasper Lippert-Rasmussen	Gary Watson
Elinor Mason	Kit Wellman
Stephen Perry	Susan Wolf

Managing Editor

David Clark

Copyeditor

Susan Wampler

Typesetting

Matthew Silverstein

SOCIAL REFORM IN A COMPLEX WORLD

Jacob Barrett

WE LIVE IN AN UNJUST WORLD. Our social and political institutions stand in need of reform. But of all the changes we might make to these institutions, which would genuinely promote justice? And how should we, as theorists, go about trying to figure this out?

Perhaps the most straightforward approach is *problem solving*: diagnosing particular problems of injustice in our world and proposing narrowly targeted institutional solutions. For example, we might aim to identify actual instances of status and resource inequality, of discrimination and oppression, of human rights violations and unjustifiable restrictions on freedom. And we might work to uncover their causes, and to come up with changes to our laws, policies, or social norms that would mitigate or eliminate them. In other words, we might adopt a relatively narrow and short-term perspective, aiming to identify “remediable injustices” in our world along with promising institutional remedies—remedies that would promote justice by chipping away at the many problems of injustice that confront us.¹

Another is *ideal theory*. On this approach, we begin not with the injustices we currently face, but by attempting to outline what the overall best institutional arrangement would be, before then figuring out “how this long-term goal might be achieved, or worked toward, usually in gradual steps.”² So instead of diagnos-

- 1 In describing the problem-solving approach I hew most closely to Wiens, “Prescribing Institutions without Ideal Theory,” but see also Sen, *The Idea of Justice* (from whom I borrow the term “remediable injustice”); Anderson, *The Imperative of Integration*, ch. 1; and Schmitz, “Nonideal Theory.”
- 2 Rawls, *The Law of Peoples*, 89. Rawls labels theorizing about how to make progress toward the ideal “nonideal theory,” but I here use the term “ideal theory” to refer to this entire approach to theorizing about reform, which includes both identifying the ideal and working out how to make progress toward it. I also set aside the interpretive question of whether Rawls aimed to identify an ideal institutional arrangement or only principles to govern such an arrangement—my concern, regardless, is with *institutional* ideal theorists who attempt the former. Of course, there are many other ways to use the term “ideal theory”; see Hamlin and Stemplowska, “Theory, Ideal Theory, and the Theory of Ideals,” and Valentini, “Ideal vs. Non-Ideal Theory.”

ing specific problems of injustice, we aim to identify the most just, problem-free institutional arrangement we could ever achieve. And instead of working out which changes to our current arrangement would mitigate or eliminate present injustice, we ask which changes would constitute progress toward this ideal. Our focus is therefore more comprehensive than on the problem-solving approach, concerned with the ideal institutional arrangement as a whole rather than with targeted solutions to particular problems. And it is also longer term, concerned not with improving justice in the short term but, most centrally, with identifying a long-term goal—an end goal—“to guide the course of social reform.”³

When we compare these two approaches side by side, ideal theory might seem alarmingly farsighted. Why would anyone think that we should focus on making progress toward a far-off goal of ideal justice, rather than on diagnosing and solving the problems of injustice we face right now? Is the whole point of theorizing about social reform not to uncover ways of ameliorating present injustice? But to this charge of hyperopia, ideal theorists retort that an exclusive focus on problem solving is itself too myopic—that even though we may solve particular problems of injustice without any ideal in mind, implementing such solutions is not, by itself, a reliable way to promote overall long-term justice, since doing so may “retard,” “stall,” or “permanently block . . . movement toward overall justice.”⁴ So when evaluating potential reforms, we cannot focus only on their short-term effect on particular problems of injustice, but must balance this against their promotion of greater justice in the long term. And while problem solving may be a good way to figure out the first half of this balance, it is not enough. We must also take into account overall long-term justice, and for that, we need ideal theory.

Ideal theorists have a point. Our world is complex—it is composed of many interacting parts—and, as I shall explain, this complexity entails that ameliorating particular problems of injustice may indeed set back the achievement of greater long-term justice. So theorizing about social reform does require more than just problem solving, as ideal theorists rightly argue. But it does not follow that we should supplement problem solving with ideal theory. In fact, I will argue that the very complexity that generates a conflict between ameliorating

3 Rawls, *A Theory of Justice*, 215.

4 Simmons, “Ideal and Nonideal Theory,” 21. Compare Buchanan, *Justice, Legitimacy, and Self-Determination*, ch. 1; Robeyns, “Ideal Theory in Theory and Practice”; and Valentini, “Ideal vs. Nonideal Theory.” Some ideal theorists go further, arguing that we cannot even address “pressing problems” of injustice without a conception of the ideal in hand (Rawls, *Theory of Justice*, 8). But this view has been subject to such a deluge of recent criticism (see Anderson, *The Imperative of Integration*, ch. 1; Sen, *The Idea of Justice*; Wiens, “Against Ideal Guidance”; and others) that it has been all but abandoned. I therefore set it aside here.

immediate problems of justice and promoting overall long-term justice also renders ideal theory epistemically overdemanding for beings like us: it makes it impossible for us either to identify the ideal or to track our progress toward it, at least with sufficient confidence to warrant its pursuit. So, thanks to complexity, problem solving is unsatisfactory and ideal theory is impracticable. The remaining question is how we ought to theorize about social reform in a complex world, and, in particular, how we should theorize about long-term justice without recourse to ideal theory.

The answer I propose is that rather than attempting to identify ways of making *progress toward* the ideal, we should instead approach questions of long-term justice by working out how to make our institutional arrangement more *progressive*: better at getting better, or more conducive to further improvements in general (though not necessarily to the achievement of any antecedently specified institutional goal). And, more concretely, I argue that the progressiveness of an institutional arrangement depends on its ability to flexibly implement many promising solutions to problems as they arise, to select for those solutions that prove successful while eliminating those that do not, and to help us learn from both our successes and our inevitable failures. On this approach, problem solving has a place, but the solutions it generates are viewed as hypotheses that function within a broader framework of institutional experimentation, selection, and learning.⁵ But ideal theory has little (if any) place: theorizing about overall long-term justice instead takes the form of figuring out how to enhance the progressiveness of this overarching framework.⁶

1. COMPLEXITY AND PROBLEM SOLVING

Let us begin with some terminology. An “institutional arrangement,” as I use the term, refers to a set of formal and informal institutions (for example, laws and

- 5 As we will see, some problem solvers endorse a similarly limited role for problem solving, and I would be attacking a straw man if I criticized them while claiming otherwise. But my purpose in this paper is not to criticize problem solving *per se*. It is to investigate the limits of problem solving and to explore how we might go beyond them.
- 6 I say “if any” because my criticism of ideal theory concerns only its ability to provide us with a long-term goal for reform, not its relevance or value in general. For all I say here, ideal theory might, for example, help us to appreciate how existing arrangements fall short of ideal justice or to uncover our basic evaluative criteria. See Swift, “The Value of Philosophy in Nonideal Circumstances”; and Gilibert, “Comparative Assessments of Justice, Political Feasibility, and Ideal Theory.” And ideal theory might also be valuable in its own right, independently of its contribution to theorizing about social reform (which is my only concern here). See Estlund, “Utopophobia.” Thanks to an anonymous reviewer for urging me to say more about this point.

social norms), as well as the background conditions that are causally relevant to their functioning (for example, facts about the natural environment, demographics, and technology). Both institutions and background conditions are what I call “institutional features”: they serve as the inputs to our (often implicit) models of how an institutional arrangement produces its effects. The features represented by the outputs of such models I call “outcomes,” and by the “justice” of an institutional arrangement I mean an evaluation of the outcome it produces given some criterion of justice. This may be an external criterion that takes into account and balances such factors as freedom, equality, oppression, and procedural justice.⁷ Or it may be an internal criterion concerned with what can be justified to actual individuals—holding, for example, that the justice of an arrangement depends on how highly it ranks on the evaluative criteria of those living under it, rather than against independently specified values or principles. Throughout, I make no attempt to defend any particular criterion of justice, but instead invite the reader to apply her favored criterion to the issues at hand. I assume only that one is not an “institutional fundamentalist” who denies that the outcome produced by an institutional arrangement is at all relevant to its justice.⁸

I emphasize the distinction between an institutional arrangement, the outcome it produces, and its justice to highlight something that, though obvious upon reflection, is too often omitted in philosophical discussions of social reform: that our evaluation of an institutional arrangement’s justice is always (logically, though perhaps not temporally) a two-step procedure.⁹ We must first map an institutional arrangement to an outcome, and only then can we map that outcome to its justice. For example, to determine whether implementing a minimum wage would improve justice, the first step is to ask what outcome this change would produce: How would it affect, say, unemployment, prices, and the distribution of income? And the second is to evaluate that outcome given our chosen criterion: Would the predicted change in such variables amount to a net increase or decrease of justice, understood, say, in Rawlsian, utilitarian, or libertarian terms? Though perhaps the bulk of political philosophy concerns the appropriate criterion to use at this second step, my focus here is the first step, where we map institutional arrangements to outcomes. For convenience, I will

7 Abstract values such as freedom and equality are sometimes themselves referred to as “ideals.” This is a perfectly fine use of language, but, to avoid ambiguity, I will never employ it myself, and will instead reserve the term “ideal” as a shorthand for “the ideally just institutional arrangement.” Thanks to an anonymous reviewer for raising this issue.

8 For a critique of such fundamentalism, see Sen, *The Idea of Justice*, ch. 3.

9 Compare Gaus, *The Tyranny of the Ideal*, ch. 2.

therefore often speak as if we can skip the second step and map institutional arrangements to justice directly. But it is important to remember what this involves. It requires us to employ a “predictive model” of how various institutional features interact to produce an outcome, which we must then evaluate in terms of its justice.¹⁰

The notion of an interaction is essential to any understanding of complexity. To say that something is complex is not merely to say that it is complicated—it is to say that it has many interacting parts.¹¹ In analyses of societies as complex systems, these “parts” are usually thought of as people, whose interactions produce emergent patterns that no individual intended or perhaps even foresaw.¹² The classic example of this is Smith’s discussion of how, given certain institutional arrangements, the market interactions of individuals each pursuing their own self-interest leads to greater social welfare.¹³ But such “invisible hand” processes are not always for the good: market interactions, to take the same example, may also lead to economic inequality, environmental destruction, and financial crises.¹⁴ And Schelling has shown how, again, given certain arrangements, individuals with mild preferences not to live in neighborhoods in which their own racial group is a small minority can lead, through a mechanism of tipping points and cascades, to stark racial segregation.¹⁵ Similar mechanisms plausibly contribute to segregation and polarization along a number of dimensions (not just race, but, for example, gender and religion) in a variety of domains (not just neighborhoods, but, for example, schools and industries).

An understanding of the various ways that individuals may unwittingly interact to produce both good and bad emergent phenomena in the presence of different institutional arrangements—which, after all, structure such interactions by constraining and incentivizing different forms of behavior—is important to any analysis of social change. In the first place, it dispels us of both the overly

10 Gaus, *The Tyranny of the Ideal*, ch. 2.

11 On complexity in general, see Mitchell, *Complexity*. For examples of its recent application to economics, see Arthur, *Complexity and the Economy*; to public policy, see Colanders and Kupers, *Complexity and the Art of Public Policy*; and to political philosophy, see Gaus, *Tyranny of the Ideal* and “The Complexity of a Diverse Moral Order.”

12 For an overview of this sort of complexity, see Miller and Page, *Complex Adaptive Systems*, pt. IV.

13 Smith, *The Wealth of Nations*.

14 These three issues are at the heart of recent attempts to rethink economic policy from a complexity perspective, especially in the wake of the 2008 financial crisis. See OECD, *Debate the Issues*.

15 Schelling, “Dynamic Models of Segregation.” Of course, much actual segregation is produced very intentionally, rather than in this way.

rationalistic view that individuals can only produce just outcomes by explicitly aiming to do so, and the overly complacent view that individuals pursuing their own projects reliably promote justice in all circumstances: both views fail to recognize that the causal relation between the achievement of justice and the intentional pursuit of other goals crucially depends on which institutions are in place.¹⁶ In the second, it helps to explain why the effects of institutional change are so difficult to predict, as it is one important source of the unanticipated consequences that often accompany such changes.¹⁷ Going forward, however, my primary concern will be not with the micro-level complexity that characterizes interactions between individuals, but instead with two forms of macro-level or institutional complexity. It will be with the way that different institutional features interact to produce outcomes, as well as to produce changes in other institutional features themselves.

Consider, first, *combinatorial complexity*. This is the sort of complexity that arises when predicting what outcome will be realized by the interaction of multiple institutional features. As is now commonplace among institutional theorists, the operation of any one institution is importantly dependent on the presence and operation of other institutions, as well as on background conditions.¹⁸ Often this phenomenon is discussed by economists under the rubric of the “general theory of second best.” As Lipsey and Lancaster famously proved, if market institutions fail to meet the set of “optimality conditions” that ensure a Pareto optimal outcome (in which no one can be made better off without someone else being made worse off), the second-best outcome is not necessarily achieved by satisfying more rather than fewer of these conditions: market institutions that fail to meet two optimality conditions might be Pareto superior (better for some and worse for none) to those that fail to meet only one.¹⁹ This, however, is just one instance of the more general phenomenon of combinatorial complexity. If we cannot have the optimal or “ideal” institutional arrangement, but the features of that arrangement interact to produce an outcome, then satisfying more of the features that compose the ideal does not necessarily result in an improvement. More generally, combinatorial complexity entails that the effect of any two institutional features cannot be reduced to the sum of the effects of each feature

16 Compare Wilson, “Two Meanings of Complex Adaptive Systems.”

17 For a classic analysis of unanticipated consequences, see Merton, “The Unanticipated Consequences of Purposive Social Action.” For a more recent overview chock-full of examples, see Tenner, *Why Things Bite Back*.

18 See especially North, *Institutions, Institutional Change and Economic Performance*.

19 Lipsey and Lancaster, “The General Theory of Second Best.”

by itself. So even when each of two institutional changes would, on their own, improve justice, both changes together might not.

Combinatorial complexity is ubiquitous. For example, market institutions only produce efficient outcomes given a background of social trust and the absence of norms prohibiting profit seeking.²⁰ Criminal prohibitions only command respect and compliance in the presence of a social norm of legal obedience, and when laws conflict too sharply with other norms this often gives rise to compliance and enforcement problems—sometimes reinforcing rather than undermining the behavior the law seeks to abolish.²¹ Color-blind policies may seem just in isolation, but may further entrench racial inequalities produced by other features of our institutional arrangement; in such cases, color-conscious policies may promote justice, even if they would undermine it given background equality.²² Or consider again the possibility that while two changes might each improve justice on their own, the combination of them might not. For example, instituting generous entitlement programs might improve justice, and opening our borders might improve justice, but doing both together might be disastrous: the influx of immigrants might result in the entitlement programs being stretched beyond the breaking point.²³

These instances of combinatorial complexity all involve features interacting to produce outcomes. But institutional features also interact in the sense that prior institutional changes may further or set back later institutional changes, and this gives rise to a type of path dependency that I will call *transitional complexity*. For example, even if both opening our borders and enacting more generous entitlement programs would indeed improve justice, opening our borders first might make it more difficult to provide more generous entitlement programs later, if (as some empirical evidence suggests) influxes of immigration undermine public support for such programs.²⁴ More general phenomena relevant to transitional complexity include lock-in and backlash. Lock-in occurs when a change prevents further changes, often because it generates interest groups who are able to maintain the new status quo.²⁵ Backlash occurs when an institutional change, say, the prohibition of alcohol or the passing of the Fugitive Slave Act, results

20 This point goes back to Smith, *The Wealth of Nations*, but see also Platteau, *Institutions, Social Norms, and Economic Development*, chs. 5, 7.

21 Gaus and I discuss this point at length in our “Laws, Norms, and Public Justification.”

22 Compare Anderson, *The Imperative of Integration*, ch. 8.

23 Related worries are common in the literature on immigration. See Carens, *The Ethics of Immigration*, ch. 12.

24 Carens, *The Ethics of Immigration*, ch. 12.

25 See North, *Institutions, Institutional Change and Economic Performance*, 94–99.

in countervailing efforts to reverse that change, or perhaps in other changes that undercut the effect of that former change.²⁶ Perhaps even more obviously, changes to secondary rules (rules for changing other rules) clearly interact with future changes to our institutions. For example, in a democracy, different voting schemes predictably lead to different policies, as do different constitutional constraints on what is subject to democratic rule.²⁷

To see the importance of combinatorial and transitional complexity, suppose that neither held. In that case, problem solving would be the perfect approach to theorizing about social reform. We could identify instances of injustice, trace their causes to particular institutional features, and identify institutional changes that would not only solve the problem we are focused on, but, in so doing, promote overall long-term justice as well. We could, for example, identify a change to employment legislation to end worker oppression, to the economic system to mitigate income inequality, to health care policy to minimize preventable deaths, to educational institutions to clear up status inequalities, to immigration policy to ameliorate global poverty, to the criminal law to mitigate racial inequality, to our gender role norms to decrease gender discrimination, and so on, without ever needing to consider how these different “solutions” would interact. We would not have to worry, for example, that the effect of different employment policies depends on what exit options employees have, which in turn depend on what sort of economic, health care, and other social safety-net programs are in place, whose cost and efficacy depend on education policies, which also affect health, criminality, the economy, norms relating to gender and race, and so on, in various crisscrossing ways, throughout the entire network of interactions. (This is combinatorial complexity.) Nor would we have to worry that changing, say, immigration and education policies would affect which further changes would occur—by, for example, generating an anti-immigrant backlash, strengthening teachers’ unions who are now able to lock us into our current system, or changing the makeup of the electorate and so influencing which further changes are likely to be democratically authorized. (This is transitional complexity.) We could, in other words, engage in problem solving without ever having to consider whether the “solutions” we generate would undermine the achievement of

26 I borrow these examples from Stuntz, “Self-Defeating Crimes.” For a general discussion of such “reactive” as opposed to “self-reinforcing” sequences, see Mahoney, “Path Dependence in Historical Sociology.” This tracks the distinction between “negative” and “positive” feedback in the literature on complexity.

27 Riker, *Liberalism against Populism*, provides an authoritative discussion of the way that different voting rules may produce different results even given the same voter preferences.

overall or long-term justice—either by interacting to make things less just in the short term or by setting back the achievement of greater justice in the long term.

But combinatorial and transitional complexity do exist. So, bringing things full circle, this is why ideal theorists are right to point out that we cannot focus only on identifying institutional changes that would ameliorate particular problems of injustice: we must also take into account how these changes would interact to produce outcomes, and to change or stabilize other institutional features.²⁸ And this suggests that we must expand our vision along two dimensions. First, we must adopt a more comprehensive, holistic attitude to evaluating institutional arrangements rather than one that focuses only on their component parts. And second, we must adopt a longer-term perspective—one that takes into account not only relatively short-term solutions to particular problems of injustice, but also whether implementing these solutions ultimately sets back or furthers future reform. Taking this suggestion to its limit, we might adopt an ideal-theoretic orientation, first trying to identify an ideally just institutional arrangement to serve as a “long-term goal of political endeavor,” and then attempting to work out how to make progress toward it.²⁹ This approach, after all, seems tailor-made to accommodating complexity: to identify an institutional arrangement as ideal, we must take into account the combinatorial interactions of all its component features, and to determine whether a short-term change constitutes progress toward this ideal, we must take into account all relevant instances of transitional complexity. But, as we will now see, this approach, too, is unsuitable for a complex world. Whereas an exclusive focus on problem solving is too myopic, we cannot see nearly as far as ideal theory presumes.

2. EPISTEMIC ASSYMETRIES AND IDEAL THEORY

The view that we ought to set the ideal as our long-term target for reform must be distinguished from the closely related view that to reform our institutional arrangements is to make them better approximate or resemble the ideal. Indeed, an appreciation of combinatorial and transitional complexity undermines whatever initial plausibility this latter view might have. If there were no combinatorial complexity, then the justice of an institutional arrangement would depend on its similarity to the ideal: every time we implemented an institutional feature that

28 Simmons appears to have combinatorial complexity in mind when he argues that “there is no reason to suppose in advance that justice in one domain is independent of justice in other domains,” and transitional complexity in mind when he warns of the potential for backlash (“Ideal and Nonideal Theory,” 21–22).

29 Rawls, *Theory of Justice*, 138.

obtains at the ideal, this would make our institutional arrangement more just. If there were no transitional complexity, then progress toward the ideal would depend on similarity to the ideal: every time we implemented an institutional feature that obtains at the ideal, we would make progress toward the ideal, since there we would be less changes left to make. But since both sorts of complexity exist, neither relation holds. Making institutional arrangements more similar to the ideal can, due to combinatorial complexity, result in a decrease in justice, or, due to transitional complexity, constitute progress away from the ideal.³⁰ And this implies that making our institutional arrangement better approximate or resemble the ideal is not a reliable way to promote justice. If ideal theory is to stand any chance of being a viable approach to theorizing about social reform, its goal cannot merely be to identify ways of making our current arrangements more similar to the ideal. Instead, it must be to identify “steps” that constitute progress toward the ideal, where making progress toward the ideal is understood in such a way that it sometimes involves making our arrangement less similar to the ideal along the way.³¹

Ultimately, I will argue that ideal theory is not, in fact, a viable approach to theorizing about social reform. But before we get to this critique, we need a better picture of ideal theory in mind. This requires us to answer three questions. First, what do we mean by the “ideal” institutional arrangement? Second, how should we cash out the idea of making “progress toward” the ideal if—as we have just seen—it does not merely amount to making our arrangement better resemble it? And third, what role is ideal theory supposed to play in a full account of social reform?

The first answer is straightforward enough. To say that something is ideal is to say that it is best, and, in this context, “best” means “most just.” But in order for the ideal to serve as a long-term goal, it is not enough for it to be the most just institutional arrangement that is (say) conceptually possible—it must be possible in the sense that there is some feasible path between it and us. Of course, this idea is very rough. The feasibility of an institutional arrangement depends both on what our world is currently like and on which ways of transforming it are compatible with various social scientific (and other) facts; but beyond this, there is ample room for disagreement. Still, no matter the details, the key point

30 This implication of transitional complexity finds no expression in recent models of the complexity of social reform, such as that found in Gaus, *Tyranny of the Ideal*, and Page, “The Imperative of Complexity.” Though such thinkers are certainly aware of transitional complexity, the model they employ assumes that making an institutional arrangement more similar to another constitutes progress toward the latter. Thanks to transitional complexity, this is not always the case.

31 Simmons, “Ideal and Nonideal Theory,” 23.

is that our judgment about whether something is feasible depends on a prediction of whether we can get there from here. To identify the ideal, we cannot simply form a conception of what a *perfectly* or *fully* just arrangement would be like given our criterion of justice, since such an arrangement might very well be infeasible, and therefore unable to play the role of a long-term goal. Instead, we must form a prediction of which institutional arrangements we could eventually realize, and then another prediction of which, of these, would produce the most just outcome—as always, given our chosen criterion.³²

Turn, then, to the second question: How should we understand “progress toward” the ideal? Well, consider the earlier worry that implementing solutions to particular problems of injustice might “retard,” “stall,” or “permanently block . . . movement toward overall justice.”³³ This worry suggests that progress should be understood in temporal terms: we make progress toward the ideal by decreasing the time it will take to get there, and progress away by increasing this time, at the limit, making it so that we will never achieve it. But since we rarely know for certain whether a change would make the ideal impossible ever to achieve (more on which shortly), in actual contexts of social reform, we typically must reformulate considerations of possibility in terms of the probability that we will ever reach the ideal, so that another way to make progress toward the ideal is to increase the probability that we will eventually get there.³⁴ This introduces a conflict internal to the notion of “progress toward”: for example, a step down a revolutionary path may have less chance of taking us to the ideal but be faster if it works, while a step down an incremental path might be more of a sure thing yet take longer. But let us set such issues aside, and assume that we have settled on a criterion that aggregates time and probability (and anything else relevant to “progress toward”) into a single standard for judging how “far” the ideal is from a particular arrangement.³⁵ In determining whether an institutional change would constitute progress toward the ideal, we must therefore form a prediction of how the resulting arrangement will continue to change. We

32 Compare Buchanan, *Justice, Legitimacy, and Self-Determination*, 61–63; and Wiens, “Political Ideals and the Feasibility Frontier.”

33 Simmons, “Ideal and Nonideal Theory,” 21.

34 Note here the structural similarity to “conditional probability” models of feasibility, on which we make something more feasible by increasing the probability we will achieve it. See Gilabert and Lawford-Smith, “Political Feasibility”; and Lawford-Smith, “Understanding Political Feasibility.”

35 For example, we might also factor in the morally relevant cost of achieving the ideal. See Rääkkä, “The Feasibility Condition in Political Theory.” But I prefer to analyze such costs separately, as something to be traded off against making progress toward the ideal rather than as an element of such progress.

must forecast forward from this initial change, asking what effect it will have on the probability that we will ever reach the ideal, the time it will take us to do so, and the extent to which it therefore constitutes progress toward or away from it.

Finally, what role does ideal theory play in a full account of social reform? Now, if identifying and pursuing the ideal is worth doing, it must be because there are some cases where implementing short-term improvements sets back progress toward the ideal. Otherwise, there would be no need to identify an ideal in order to make progress toward it: every short-term improvement would simultaneously constitute progress toward the ideal, so we could make such progress simply by implementing the short-term improvements we discover through problem solving.³⁶ But, as we have seen, complexity does generate a trade-off between short-term justice and progress toward the ideal, so there are indeed cases where pursuing the ideal comes at the expense of ameliorating present injustice. This, however, is not to say that we should care only about progress toward the ideal, and we should not saddle ideal theorists with such an extreme commitment: as they emphasize, there are times when ignoring short-term injustice in order to make progress toward the ideal is morally impermissible or otherwise not worth the cost.³⁷ So we should understand ideal theorists as claiming not that we ought always to pursue the ideal at the expense of short-term justice, but that we ought to do so sometimes—at least in some nontrivial range of cases where the expected long-term benefit of pursuing the ideal outweighs the expected short-term cost of forgoing a short-term improvement.

So understood, ideal theory is both maximally comprehensive and maximally long term, and it might therefore seem to fully accommodate the complexity of our world. But, alas, the very complexity that makes ideal theory attractive also makes it impracticable—at least for agents like us. For combinatorial and transitional complexity not only ensure that short-term solutions to particular problems of injustice sometimes conflict with progress toward the ideal, they also give rise to two epistemic asymmetries. First, due to combinatorial complexity, as we consider larger changes to (more features of) our institutional arrangement, our predictions about what outcomes those changes will produce, and therefore about what their effects on justice will be, decrease in reliability. Second, due to transitional complexity, as we forecast the effects of institutional changes further into the future, our predictions about which subsequent changes will occur become less reliable as well. And these two asymmetries undermine the epistemic presumptions of ideal theory. We cannot identify the ideal institutional arrangement with sufficient confidence to warrant pursuing

36 Gaus, *Tyranny of the Ideal*, 82–84.

37 Rawls, *The Law of Peoples*, 89; Simmons, “Ideal and Nonideal Theory,” 18–22.

it at the expense of short-term justice. And even if we could, we would still lack the epistemic wherewithal to identify changes to our current arrangement that would constitute progress toward it with this requisite degree of confidence.

Consider first a recent argument of Gaus's, which begins by noting that our predictive models of how institutional arrangements interact to produce outcomes are not very accurate in the first place—they not only come with a probabilistic margin of error, but may also fail to assign any probabilities to wholly unanticipated consequences.³⁸ Furthermore, like all models of complex systems, they are subject to “error inflation.”³⁹ We may calibrate our models of actual institutional arrangement to the data: if we predict, say, that increasing the minimum wage will spike unemployment, or that lowering the corporate tax rate will increase inequality, but find that this does not occur—or that either change produces some wholly unanticipated effect, say, on the gendered or racial division of labor—we may go back and revise our model in light of this feedback. But when it comes to models of merely hypothetical arrangements, we cannot calibrate our models in this way, and so are more prone to error. Error inflation then occurs as we consider institutional arrangements that differ more and more in their institutional features from actual ones. Gaus explains: “An error in predicting the workings of one feature will spread to errors in predicting the justice-relevant workings of interconnected features, magnifying the original error. As this new erroneous model is used as the basis for understanding yet further arrangements, the magnified errors become part of the new model, which is then itself subject to the same dynamic.”⁴⁰

The upshot of error inflation is clear. We should have more confidence in our prediction of the effect of a change, say, to either the minimum wage or the corporate tax rate, than in our prediction of a change to both, since the errors we make in predicting the effect of each carry over into our prediction of how they interact. We should have more confidence in this than in our prediction of how a radically redesigned economic system such as market socialism or property-owning democracy would work, since such systems differ from actual ones in so many ways that errors massively inflate. And we should have even less confidence—indeed, basically none at all—in our ability to predict the outcome produced by an even more divergent arrangement, designed to handle not only economic injustice, but racial and gender injustice, global injustice, and all other

38 These two possibilities track Knight's distinction between probabilistic “risk” and non-probabilistic “uncertainty” (*Risk, Uncertainty, and Profit*).

39 Gaus, *Tyranny of the Ideal*, 80. On error inflation more generally, see Smith, *Explaining Chaos*.

40 Gaus, *Tyranny of the Ideal*, 80.

forms of injustice as well. So, since most candidate ideals differ radically from any actual institutional arrangement, we lack the ability to make a confident prediction about the outcome such arrangements will produce, and, therefore, to evaluate their justice. An essential presupposition of ideal theory cannot be met: we cannot judge which institutional arrangement is ideal, and therefore worth pursuing at the expense of short-term justice, with sufficient confidence to license this pursuit.

Gaus's epistemic critique of ideal theory appeals only to combinatorial complexity, but in case one is not yet convinced, we may considerably strengthen it by appealing to transitional complexity as well. To begin, note that, as a perfectly general matter, our ability to predict the future becomes less reliable as we attempt to forecast further in time. The relevant mechanism here is once again error inflation: the errors we make in predicting what will happen tomorrow get carried over into our prediction of what will happen the day after, which get carried over into our prediction of what will happen next week, next year, next decade, and so on. And this general tendency is magnified in complex systems, where, due to transitional complexity, it becomes impossible for us to predict anything in the very long term. For, in the first place, doing so requires us to predict where backlash will occur, where we will get locked in, and, more generally, how people will respond to changes to our institutional arrangement by producing further changes, and others to those changes, and so on, far into the future. But these predictions are notoriously difficult to make, not only because each depends on our prior prediction, but also because predicting individuals' responses to institutional changes requires us to predict what outcome those changes will realize. And, in the second, this requires us to predict the occasion and effect of technological innovations and external shocks. Yet it is deeply implausible to think that we could have predicted the occasion of or institutional change caused by the invention of the printing press, telephone, radio, television, or Internet, or by the occurrence of the industrial revolution, either world war, or, going forward, climate change. And it is similarly implausible to think that we can reliably forecast which further technological changes or external shocks will occur.

As a result, as we attempt to forecast further into the future, not only do our probabilistic judgments of how individuals will respond to changes multiply together and decrease our confidence in which further changes will occur, but the probability of totally unexpected events increases as well. The upshot is that it is impossible for us to forecast institutional change far into the future at all: we cannot determine which institutional arrangements we could get to in the very

long term, nor which short-term changes bring us toward them, at least with any reasonable degree of confidence.⁴¹

On the basis of similar considerations, Wiens concludes that “given the number of variables to which our feasibility assessments must be sensitive, the complexity of their interactions, and the potential for path-dependence, determining whether any particular long-range objective is feasible is beyond human cognitive capacity.”⁴² In other words, whereas Gaus worries that we cannot figure out how just various candidate ideals would be, Wiens worries that “we simply cannot determine with any confidence whether particular long-range objectives are feasible, let alone with sufficient confidence to justify adopting a political ideal as a reform target.”⁴³ And since, as I have emphasized, identifying the ideal requires us to figure out both of these things—to determine which *feasible* institutional arrangement is *most just*—this suggests that ideal theory is an impossible enterprise.⁴⁴ Indeed, the problem is even worse than Gaus or Wiens suggests, since even if we somehow knew which institutional arrangement was ideal, we would still lack the ability to confidently identify which short-term changes would constitute progress toward it. To do so, we would again have to forecast whether such changes would increase the probability of us ever achieving the very long-term goal of reaching the ideal or decrease the time it would take to do so. But, as we have just seen, we cannot confidently forecast the effect of institutional changes far into the future this way. Backlash alone illustrates the problem, since backlash can result in changes that appear to be going in one direction actually causing the reverse, and because predicting backlash in part requires us to predict what outcomes will be produced—something we cannot do when it comes to institutional arrangements that are very dissimilar to our own. And once we factor in considerations of lock-in, technological change, external shocks, and so on, our epistemic situation looks even more hopeless.

Summing up our discussion, then, ideal theory requires us to confidently determine three things: which institutional arrangements are feasible, which of these is most just, and which short-term changes constitute progress toward this ideal. But, thanks to complexity, we cannot perform any of these tasks—

41 Compare Hayek, *The Constitution of Liberty*, ch. 2; and Tetlock, *Expert Political Judgment*.

42 Wiens, “Political Ideals and the Feasibility Frontier,” 467.

43 Wiens, “Political Ideals and the Feasibility Frontier,” 467.

44 As Buchanan argues, if a conception of the ideal is to serve as an appropriate long-term goal for reform, it must be not only *causally* feasible but also *morally* feasible in the sense that it can be reached through morally permissible means (*Justice, Legitimacy, and Self-Determination*, 61–63). This only adds to the difficulty of determining which institutional arrangements are feasible, and, therefore, which is ideal.

let alone all three—so we cannot identify steps that would constitute progress toward the ideal with sufficient confidence to warrant taking these steps at the expense of short-term justice. Though focusing only on identifying narrow solutions to particular problems of injustice is not a reliable way to promote overall long-term justice, trying instead to figure out how to make progress toward the ideal is beyond our epistemic ken. The trick will be to see if we can thread the needle, and avoid both the myopia of problem solving and the epistemic over-demandingness of ideal theory.⁴⁵

3. EXPERIMENTATION AND PROGRESSIVENESS

So far, we have seen that, in theorizing about social reform, we face two competing pressures. Because of the potential for interaction between different narrow solutions to immediate problems of injustice, there is a pressure to expand our sights outward and forward toward more comprehensive changes and their longer-term effects. But at the same time, our ability to predict the effect of institutional change rapidly deteriorates as we attempt to expand our sights in either of these ways. So there is a contravening pressure to contract our focus back to the narrow and the short term.

Now, to be clear, the problem raised by these competing pressures is not just one for ideal theorists, nor is it one that depends on the precise limits of our predictive capacities. Even if we were fairly adept at predicting the effects of and tracking our progress toward, say, medium-sized, medium-term changes, there would remain a gap between the largest and longest-term reforms whose effects we could confidently identify and pursue, and more comprehensive institutional changes that would produce greater justice in the long term. And implementing these medium-sized, medium-term changes would still risk setting back the pursuit of overall long-term justice, by interacting either to undermine overall justice (due to combinatorial complexity) or to set back future improvements (due to transitional complexity), just like narrow, short-term changes. This leaves us with our central methodological challenge: *How can we identify institutional reforms that we can predict with reasonable confidence will promote overall long-term*

45 One further worry for ideal theorists is that there may be no fixed ideal institutional arrangement to pursue in the first place, since, due to combinatorial complexity, “institutions adopted for a particular time, even if optimal . . . at that time, may be far from optimal as the human environment changes over time” (North, *Understanding the Process of Economic Change*, 132). I lack the space to develop this “moving target” objection here, but see Muldoon, *Social Contract Theory for a Diverse World*, 6, 29; Popper, *The Open Society and Its Enemies*, 174; and Rosenberg, “On the Very Idea of Ideal Theory in Political Philosophy,” 64–70. Thanks to an anonymous reviewer for suggesting I flag this worry here.

justice, when we cannot figure out which institutional arrangements to aim at in the long term, or how to make progress toward them?

Although problem solvers rarely address this question directly, they sometimes appear to suggest that we cannot meet this challenge, and that we should therefore scale back our ambitions and carry on with problem solving while taking into account combinatorial and transitional complexity to the limited extent we can. For example, Wiens writes that when we engage in problem solving, we should “avoid negative interactions as far as possible” and do our best to “keep open possibilities for future improvement.”⁴⁶ But since we are not very good at this, we must recognize that the solutions we generate are “tentative and experimental.”⁴⁷ We must give up on anything approaching certainty and, as Anderson puts it, treat “imagined solutions to identified problems . . . as *hypotheses*, to be tested in experience.”⁴⁸

Problem solvers who endorse this experimental orientation are on the right track, and we are now well positioned to understand why. In theorizing about social reform, we face two epistemic asymmetries: we are worse at predicting the effects of larger changes than smaller changes, and at predicting the longer-term effects of changes than their shorter-term effects. Indeed, at the limit, we are much worse at predicting the effect of any change than we are at evaluating our current arrangement, since every institutional change brings with it some probabilistic margin of error and some risk of totally unanticipated consequences. Thankfully, these asymmetries have a flip side. We are better at evaluating institutional changes after they are implemented than we are at predicting what they will do, better at evaluating where we have ended up than predicting where we will go. And this is why the ultimate test of any proposed solution’s effect on justice must be how it works out in practice. Problem solving can at best serve as a means of hypothesis generation. Its goal must be to discover institutional reforms that are “worth a try,”⁴⁹ but whose actual effects on justice can only be ascertained through trying them out in various combinations—that is, through social experimentation.⁵⁰

Once we recognize our epistemic limitations and the resultant need to test

46 Wiens, “Prescribing Institutions without Ideal Theory,” 66.

47 Wiens, “Prescribing Institutions without Ideal Theory,” 66.

48 Anderson, *The Imperative of Integration*, 6.

49 Schmidt, “A Realistic Political Ideal,” 772.

50 An emphasis on social experimentation as a means for promoting reform is nothing new, but is a running theme throughout the history of political philosophy. See especially Mill, *On Liberty*; Dewey, *The Public and Its Problems*; and Popper, *The Open Society and Its Enemies*.

our proposals in practice, however, this mandates a far more significant shift in our approach to theorizing about social reform than problem solvers seem to realize. To see why, let us consider the question of what makes a proposed solution worth a try. One factor is the effect we predict it will have on problems of injustice, taking into account complexity to whatever extent we can. But that is not all. For one thing, it also matters how much experiments teach us about how to improve justice in the future, since some experiments provide more useful data than others that we may feed back into our causal models to develop more promising institutional solutions going forward. Most obviously, we learn more from novel institutional experiments than from those we are already familiar with: implementing such solutions allows us to generate new data by *exploring* the space of institutional possibilities rather than merely *exploiting* our current knowledge of what has worked in the past.⁵¹ Perhaps less obviously, we must also take into account the epistemic quality of the experiments we engage in. For example, more radical experiments generally have less internal validity than more modest ones, since their effects depend on the interactions between so many variables that it is difficult to trace out the causation. But there is also the worry that small-scale social experiments lack external validity: that their effects will not “scale up” to the societal level.⁵² Thankfully, these are not mutually exclusive alternatives: we may also engage in relatively modest experiments at relatively large scales, arguably maintaining a reasonable level of internal and external validity.⁵³ Of course, there is much room for further debate here concerning the epistemic merits of different types of social experiments, and there is already a large literature on the subject.⁵⁴ But the general point is that, in thinking about which experiments are worth a try, we must consider not only their predicted effects on justice, but also what we expect to learn from them. Sometimes, the better experiment may not be the one that we predict to be more just in the short term, but the one that will teach us more going forward.

This might seem more like mad science than social reform. Why should we forgo ameliorating present injustice in order to gather social scientific data? The answer is that we face a trade-off between short-term and long-term justice, and that the better models we have of how institutional features interact to produce

51 March, “Exploration and Exploitation in Organization Learning,” provides a classic discussion of the exploration/exploitation trade-off. See also D’Agostino, *Naturalizing Epistemology*.

52 Gaus, *Tyranny of the Ideal*, 89–93.

53 Popper, *The Open Society and Its Enemies*, 176.

54 For a start, see Shadish, Cook, and Campbell, *Experimental and Quasi-Experimental Design for Generalized Causal Inference*; and Cartwright and Hardie, *Evidence-Based Policy*.

outcomes, the better positioned we are to promote justice in the future. Of course, we should not always forgo short-term improvements for this reason, maniacally expanding our knowledge without ever putting it to use. We should sometimes resolve the trade-off in favor of short-term justice, and we should certainly refrain from experiments that are morally impermissible—say, because they impose severe risks on individuals who do not voluntarily bear them, or because their costs systematically fall on already disadvantaged groups.⁵⁵ But just as ideal theorists argue that we must sometimes forgo short-term justice in order to make progress toward the ideal, my suggestion is that we must sometimes forgo short-term justice in order to better position ourselves to promote further justice. We must trade off the predicted short-term justice produced by a social change not against how much it constitutes *progress toward* the ideal, but against how *progressive* it is: how conducive it is to further improvements in general, though not necessarily to the achievement of any antecedently specified goal. And one factor that is relevant to the progressiveness of an institutional arrangement is how well we understand how to improve justice from there.

This brings us to another way of theorizing about long-term justice that does not qualify as either problem solving or ideal theory. In particular, while improving our understanding of how institutional features interact is one way for us to improve our progressiveness, we can similarly improve our progressiveness by enhancing the framework within which experimentation takes place—by making it more amenable to learning. In part, we might do so by improving individual epistemic abilities. But at the institutional level, we might also improve the social epistemic conditions in which we theorize about reform. To some extent, this depends on the existence of social epistemic conditions that are conducive to good inquiry in general, including free speech, a diversity of research agendas, shared vocabularies, and so on, that philosophers of science, at least since Kuhn, have made much progress exploring.⁵⁶ But it also depends more specifically on the extent to which we have mechanisms in place for monitoring the institutional experiments we engage in, gathering information from the experiences of other past and present societies, and storing it in our institutional memory.

Putting this idea more generally: just as we earlier understood making *progress toward* the ideal as increasing the probability that we will achieve the ideal

55 Müller, *Political Pluralism, Disagreement, and Justice*, 176–77; Knight and Johnson, *The Priority of Democracy*, 48–49.

56 Kuhn, *The Structure of Scientific Revolutions*. See also Kitcher, *The Advancement of Science*; D'Agostino, *Naturalizing Epistemology*; and, for a discussion of social epistemic conditions that are conducive to good moral inquiry in particular, Buchanan and Powell, *The Evolution of Moral Progress*.

or decreasing the time it will take to do so, we may now similarly understand improving the *progressiveness* of an institutional arrangement as increasing the probability and speed with which it will continue to improve justice indefinitely into the future. But the progressiveness of our institutional arrangement depends not only on its amenability to learning; it also depends on the extent to which our arrangement permits flexible experimentation going forward. Among other things, this requires the replacement of norms of dogmatism with those permitting experimentation, the ability to avoid lock-in due to seizure by interest groups, and a reluctance to implement policies that are difficult to reverse.⁵⁷ But since our goal is to promote justice, it would be too simplistic to think that we always ought to avoid lock-in, that we always want to leave all options open going forward. Sometimes, we do want to close options off, at least temporarily: we want institutions that genuinely solve problems of injustice to remain stable as long as they remain solutions. And this mandates a reliance not only on learning mechanisms that help us to update our causal models in response to feedback about the effect of our experiments, but also on selection mechanisms by which we can stabilize successful experiments and eliminate failed ones.

To unpack this idea further, let us say that an institutional feature is “worth keeping” if, taking into account all the practical and epistemic benefits and costs it provides, there is no feasible change to it that is “worth a try.” And let us say that an institutional experiment is a “success” if it is worth keeping, a “failure” if it is not. Now, the reason we have to engage in institutional experimentation in the first place is that we cannot confidently predict which institutional changes will be successes and failures. And this difficulty is only magnified by the fact that, due to combinatorial complexity, a feature that is worth keeping at one time may cease to be worth keeping at another, after other changes have occurred that interact with it. So we often cannot predict which of the institutional features we implement will prove worth keeping, nor how long they will remain that way—especially as we attempt to forecast the effects of these changes further into the future. What we need, then, are selection mechanisms that allow us to stabilize institutional features that prove worth keeping and to modify those that are not. Progressiveness, in other words, depends not only on learning mechanisms of epistemic feedback, but on selection mechanisms of practical feedback as well.

Engaging in a wide range of social experimentation against the backdrop of

57 The need to change our informal norms in this way is a running theme of both Dewey, *The Public and Its Problems*, and Popper, *The Open Society and Its Enemies*. For more on preventing lock-in due to seizure by interest groups and avoiding irreversible policies, see, respectively, North, *Understanding the Process of Economic Change*, 125; and Campbell, “Reforms as Experiments,” 410.

an institutional framework that embodies mechanisms of epistemic and practical feedback—of learning and selection—is the only way that agents like us can promote long-term justice. Due to combinatorial complexity and the epistemic asymmetry to which it gives rise, we cannot hope to predict the effects of large-scale, long-term changes (or the sequential combinations of smaller changes). So the only way for us to determine the effect of combinations of institutional features is to try them out and monitor them after the fact. Due to transitional complexity and the epistemic asymmetry it produces, we cannot hope to predict what the effect of short-term changes will be on the institutional arrangement we will end up with in the long run. So the only way for us to pursue long-term reform is through mechanisms that allow for continual adjustments of our institutional arrangement on the basis of the feedback we gather from this experimentation. Thus, whereas experimentation and learning mechanisms of epistemic feedback allow us to tame the epistemic difficulties raised by combinatorial complexity by expanding our predictive capacities and reducing the error that goes into such predictions, the goal of selection mechanisms of practical feedback is not to tame but to harness transitional complexity by reducing our reliance on prediction through the correction of errors after the fact. Though we cannot predict where phenomena like backlash and lock-in will occur, the hope is that we can institutionalize mechanisms that correlate backlash (or a functional equivalent) with institutional features that prove worth changing, and lock-in (or a functional equivalent) with those that prove not to be. That is what selection mechanisms are meant to do.

This is all rather abstract. To make it more concrete, let us examine two approaches to realizing these mechanisms currently popular in the literature. The first is experimental democracy.⁵⁸ Here, the rough idea is that experimentation is achieved through democratic deliberation and voting on which reforms to implement, and selection and learning occur through the monitoring of these institutions' effects and then deliberating and voting on which institutions to maintain and which to eliminate. Thus, experimentation is achieved primarily through democratically authorized "reforms as experiments," and selection primarily through deliberating and voting on which experiments to maintain, and we learn primarily by gathering evidence about the effects of these various reforms.⁵⁹ Long-term progressiveness is therefore achieved via a sort of centralized democratic experimenter, as opposed to the more canonical central planner

58 For historical defenses of experimental democracy, see Dewey, *The Public and Its Problems*; and Popper, *The Open Society and Its Enemies*. For more recent defenses, see Anderson, "The Epistemology of Democracy"; Knight and Johnson, *The Priority of Democracy*.

59 Campbell, "Reforms as Experiments."

on the one hand, or a decentralized mechanism on the other. And such a system is able to improve its own progressiveness over time as it applies this method reflexively—to the very features that provide for experimentation, selection, and learning.⁶⁰

This brings us to the second major approach: polycentricity.⁶¹ This time, the rough idea is that experimentation is achieved not only through consecutive or diachronic experimentation, but also through a number of institutions being tried out simultaneously in different jurisdictions. So, for example, in a federalist system, there are a number of distinct political jurisdictions that, though bound together by common federal laws, have decision-making authority over a range of issues within their territory. Or, at the informal level, different social groups, though bound together by common laws or norms, may simultaneously try out different informal norms over a range of issues where their shared institutions are silent.⁶² In each case, experimentation is achieved through different groups employing different laws, policies, or norms at the same time; selection occurs as groups or rules compete for adherents; and learning arises through groups monitoring the results not only of their own (formal or informal) institutions, but also those of others, and adjusting their own institutions accordingly.

We need not settle here the debate between experimental democrats and polycentrists. Indeed, the conflict between the two approaches is less stark than I have just made it seem. Experimental democrats, for example, generally recognize the importance of some degree of polycentricity, as well as the necessity of not only formal governmental procedures, but also such feedback mechanisms “as periodic elections, a vigilant press, petitions to government, and public commentary on proposed administrative regulations” as well as “disruptive demonstrations and legal action.”⁶³ And polycentrists universally recognize the importance of an overarching (typically democratic) governance structure to oversee the experiments that take place at its various centers of decision making, facilitating information sharing and minimizing negative externalities. The difference between experimental democrats and polycentrists is therefore more a matter of emphasis than anything else. It is best understood as a disagreement over the

60 Knight and Johnson, *The Priority of Democracy*, ch. 6.

61 Here, we find an approach primarily advocated for by political economists such as Ostrom (*Understanding Institutional Diversity*) and Aligica and Tarko (“Polycentricity”), but also more recently by philosophers such as Müller (*Political Pluralism, Disagreement, and Justice*) and Gaus (*Tyranny of the Ideal*, ch. 4).

62 Gaus, *Tyranny of the Ideal*, ch. 4.

63 Anderson, *The Imperative of Integration*, 99.

extent to which progressiveness depends on centralized or decentralized, formal or informal processes—not over whether such processes matter at all.

In any event, I mention these approaches here only to render more concrete what a reasonable level of progressiveness would look like in practice, as well as the sort of questions we must ask about progressiveness more generally. In addition to questions concerning the extent to which progressiveness depends on centralized or decentralized, formal or informal processes, we must also ask about the relative importance of learning and selection mechanisms: Do we improve progressiveness primarily by getting better at predicting what will work, or by getting better at stabilizing what has worked and eliminating what has not? Presumably, those who are more optimistic about our predictive capacities will emphasize the former, while those more pessimistic will emphasize the latter—at the limit, abandoning prediction altogether in favor of a pure evolutionary mechanism of random variation and selection.⁶⁴ Another salient topic concerns the role of moral diversity in progressive arrangements. Throughout this paper I have simply set aside the issue of what criterion of justice to use when evaluating institutional arrangements by the outcomes they produce. But there is much disagreement about such matters in the real world, and this raises a number of further questions about progressiveness. For example, do progressive institutions require that their members at least form an “overlapping consensus” on a reasonable “political conception of justice”?⁶⁵ Or can a diversity of moral and political views—including those disagreeing about what criterion to use when

64 The extent to which we can predict the effects of institutional change, and improve these predictions through learning, depends largely on just how complex our world is. An extreme view is that our world is chaotic: the interactions between its features are so dense that minor tweaks to institutions reverberate throughout the entire arrangement in entirely unpredictable ways (“the butterfly effect”). But the relative stability of our world, the history of successful social reform, and the fact that our understanding of institutional functioning has clearly improved across time all suggest that this is not so—that our world, though complex, is “nearly decomposable,” such that “the short-run behavior of each [institutional feature] is approximately independent of the short-run behavior of the other components” (Simon, “The Architecture of Complexity,” 474; compare Buchanan and Powell, *The Evolution of Moral Progress*, 263–64). Put in these terms, the debate over our predictive capabilities turns on just how “approximate” this independence is, as well as on how adept we are at identifying the boundaries of approximately independent features. Thankfully, I need not resolve this debate here, since the argument of this paper requires only that features are not so independent that we can confidently identify and track our progress toward the ideal, but not so interdependent that intentional social reform is entirely beyond our ken.

65 Rawls, *Political Liberalism*.

evaluating the outcomes produced by our institutions—coexist with progressive institutions, or even be a driver of progress, as some have recently argued?⁶⁶

Again, I cannot hope to answer such questions here, but only to put them on the agenda. Still, the very fact that they appear to admit a wide range of answers suggests an obvious objection. If it is epistemically infeasible for us to confidently determine what would constitute progress toward the ideally just institutional arrangement, then why is it not similarly infeasible for us to determine what would constitute progress toward an ideally progressive arrangement? Why is my own approach not just as epistemically overdemanding as ideal theory?

Answering this objection provides me with an opportunity to clarify my position. My claim is not that we must identify the ideally progressive institutional arrangement so that we can trade off short-term improvements in justice against progress toward this progressive ideal. It is rather that we must identify short-term improvements in progressiveness, so that we can trade off short-term improvements in justice against short-term improvements in progressiveness. In so doing, we avoid the epistemic excesses of ideal theory, because we may adopt the same orientation as problem solvers: aiming to identify and solve problems that undermine not justice, but progressiveness. For example, we might attempt to identify and mitigate biases that feed into our current selection or learning mechanisms—such as the tendency of institutions to change in ways that favor the short-term interests of the rich and powerful rather than long-term justice, or the fact that when we monitor our existing arrangement we often give undue weight to the opinions of some rather than others.⁶⁷ Similarly, we might attempt to solve incentive problems that make our institutional arrangement less conducive to learning and selection—for example, the fact that in most democracies, politicians have an incentive to oversell the benefits of their proposed reforms (to increase their chance of being passed), and then a further incentive to prevent the monitoring of these reforms (because they are unlikely to live up to their bill of goods).⁶⁸ Or we might attempt to devise ways to avoid dogmatism and resistance to experimentalism more generally, to avoid repeating the mistakes of the past through improved institutional memory and cross-jurisdictional information sharing, and so on.

Of course, there remains the risk that, in only considering short-term justice and progressiveness in this way, we ultimately make progress away from the ide-

66 See, for example, Gaus, *Tyranny of the Ideal*; Muldoon, *Social Contract Theory for a Diverse World*; and Müller, *Political Pluralism, Disagreement, and Justice*.

67 This is a driving concern of Knight and Johnson, *The Priority of Democracy*. See Fricker, *Epistemic Injustice*.

68 Campbell, "Reforms as Experiments," 410.

ally just or progressive institutional arrangement. But the severity of this problem is mitigated in two ways. First, the risk itself is considerably less than in the case of pursuing ideal justice, given that short-term improvements in progressiveness generally enhance our understanding of institutional functioning and our ability to make further positive changes, and so generally further rather than set back our ability to improve both justice and progressiveness. And, second, I do not rest my argument on the claim that changing our institutional arrangement in ways we predict will improve progressiveness is in all cases guaranteed to maximally promote long-term justice. No such guarantee is available. Instead, I rest it on the comparative claim that trading off short-term improvements in justice against short-term improvements in progressiveness is a better approach to social reform than either focusing only on short-term justice or attempting to trade off short-term justice against progress toward ideal justice. It represents the appropriate middle ground between the myopia of the former approach and the impracticability of the latter.

4. CONCLUSION

Drawing all these threads together, there are, on the approach I have outlined, two basic tasks for theorists of social reform.⁶⁹ The first is to engage in problem solving: attempting to identify, to the best of our ability, institutional changes that would ameliorate particular instances of injustice. But, given the complexity of our world and the epistemic limitations it generates, we must recognize that we cannot really come up with surefire solutions, so much as hypotheses that are worth a try. And in evaluating such hypotheses, we must take into account not only the extent to which we predict they will solve such problems, but also the extent to which they affect our progressiveness or prospects for future reform going forward—for example, how much we will learn from them, and how difficult they will be to reverse. Or, to take a different sort of example, if we live in a democracy whose mechanisms of experimentation, learning, and selection depend on public trust and participation in the democratic process, then we must count it against an unpopular policy that passing it would lead to public distrust or alienation from the process—even if the policy would genuinely ameliorate injustice in the short term, its popularity notwithstanding.

69 That there are only two basic tasks is compatible with there being other subsidiary tasks that inform these basic ones. For example, one such task may be pure normative theorizing about the appropriate criterion of justice to use when evaluating outcomes. For illustrations of the sort of theorizing I have in mind, see Barrett, “Is Maximin Egalitarian?” and “Efficient Inequalities.”

The second basic task is to theorize directly about how to improve the progressiveness of our institutional arrangement: the speed and reliability with which it will continue to improve in justice. This, I have argued, depends on its conduciveness to a wide range of promising experiments, to selecting for successful ones, and to learning from both successes and failures. Theorizing about progressiveness differs from problem solving because it is not concerned with ameliorating particular problems of injustice, but rather with improving the progressiveness of the framework within which such problem solving occurs. So while ideal theorists are right that problem solving is not enough, they are wrong that we need to supplement problem solving with ideal theory. Instead, we must supplement problem solving with theorizing about how to make our institutional arrangement more progressive. And we must trade off improvements in short-term justice not against progress toward the ideal, but against progressiveness more generally.

The various questions I have flagged about progressiveness are difficult questions with no easy answers—and we have only scratched the surface of the many issues that progressiveness raises. But it is precisely these issues to which theorists of social reform must now turn. In a complex world, we cannot assume that ameliorating particular instances of injustice promotes greater long-term justice, but neither can we identify or track our progress toward a long-term goal of ideal justice. So it is only by identifying institutional changes that improve progressiveness that we can figure out how to promote long-term justice. And it is only by implementing such changes that we can effectively pursue it.⁷⁰

University of Arizona
jacobbarrett@email.arizona.edu

REFERENCES

- Aligica, Paul D., and Vlad Tarko. "Polycentricity: From Polanyi to Ostrom, and Beyond." *Governance* 25, no. 2 (September 2011): 237–62.
- Anderson, Elizabeth. "The Epistemology of Democracy." *Episteme* 3, nos. 1–2 (June 2006): 8–22.

70 I would like to thank Allen Buchanan, Tom Christiano, Chris Howard, Brian Kogelmann, Sarah Raskoff, Connie Rosati, Santiago Sanchez Borboa, David Schmitz, two anonymous reviewers, and especially Jerry Gaus for helpful comments and discussion. Thanks also to participants at the 2018 PPE Society Conference in New Orleans, as well as to the members of the University of Arizona Philosophy Department's "Works in Progress" group and of Allen Buchanan's spring 2019 graduate seminar on "Evolution and Large-Scale Moral Change."

- . *The Imperative of Integration*. Princeton, NJ: Princeton University Press, 2010.
- Arthur, W. Brian. *Complexity and the Economy*. New York: Oxford University Press, 2014.
- Barrett, Jacob. “Efficient Inequalities.” *Journal of Political Philosophy*, forthcoming.
- . “Is Maximin Egalitarian?” *Synthese* 197, no. 2 (February 2020): 817–37.
- Barrett, Jacob, and Gerald Gaus. “Laws, Norms, and Public Justification: The Limits of Law as an Instrument of Reform.” In *Public Reason and Courts*, edited by Silje Aambø Langvatn, Wojciech Sadurski, and Mattias Kumm, 201–28. Cambridge: Cambridge University Press, forthcoming.
- Buchanan, Allen. *Justice, Legitimacy, and Self-Determination: Moral Foundations for International Law*. New York: Oxford University Press, 2004.
- Buchanan, Allen, and Russell Powell. *The Evolution of Moral Progress: A Biocultural Theory*. New York: Oxford University Press, 2018.
- Campbell, Donald T. “Reforms as Experiments.” *American Psychologist* 24, no. 4 (April 1969): 409–29.
- Carens, Joseph. *The Ethics of Immigration*. New York: Oxford University Press, 2013.
- Cartwright, Nancy, and Jeremy Hardie. *Evidence-Based Policy: A Practical Guide to Doing It Better*. New York: Oxford University Press, 2012.
- Colanders, David, and Ronald Kupers. *Complexity and the Art of Public Policy: Solving Society’s Problems from the Bottom Up*. Princeton, NJ: Princeton University Press, 2016.
- D’Agostino, Fred. *Naturalizing Epistemology: Thomas Kuhn and the ‘Essential Tension.’* New York: Palgrave Macmillan, 2009.
- Dewey, John. *The Public and Its Problems*. New York: Henry Holt and Co., 1927.
- Estlund, David. “Utopophobia.” *Philosophy and Public Affairs* 42, no. 2 (August 2014): 207–35.
- Fricker, Miranda. *Epistemic Injustice: Power and the Ethics of Knowing*. New York: Oxford University Press, 2007.
- Gaus, Gerald. “The Complexity of a Diverse Moral Order.” *The Georgetown Journal of Law and Public Policy* 16 (Fall 2018): 645–80.
- . *Tyranny of the Ideal: Justice in a Diverse Society*. Princeton, NJ: Princeton University Press, 2016.
- Gilbert, Pablo. “Comparative Assessments of Justice, Political Feasibility, and Ideal Theory.” *Ethical Theory and Moral Practice* 15, no. 1 (February 2012): 39–56.

- Gilbert, Pablo, and Holly Lawford-Smith. "Political Feasibility: A Conceptual Exploration." *Political Studies* 60, no. 4 (February 2012): 809–25.
- Hamlin, Alan, and Zofia Stemplowska. "Theory, Ideal Theory, and the Theory of Ideals." *Political Studies Review* 10, no. 1 (January 2012): 48–62.
- Hayek, F.A. *The Constitution of Liberty*. Chicago: University of Chicago Press, 1960.
- Kitcher, Philip. *The Advancement of Science: Science without Legend, Objectivity without Illusions*. New York: Oxford University Press, 1993.
- Knight, Frank H. *Risk, Uncertainty, and Profit*. Boston: Houghton Mifflin Company, 1921.
- Knight, Jack, and James Johnson. *The Priority of Democracy: Political Consequences of Pragmatism*. Princeton, NJ: Princeton University Press, 2011.
- Kuhn, Thomas. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.
- Lawford-Smith, Holly. "Understanding Political Feasibility." *Journal of Political Philosophy* 21, no. 3 (September 2013): 243–59.
- Lipsey, R. G., and Kevin Lancaster. "The General Theory of the Second Best." *The Review of Economic Studies* 24, no. 1 (1956–57): 11–32.
- Mahoney, James. "Path Dependence in Historical Sociology." *Theory and Society* 29, no. 4 (August 2000): 507–48.
- March, James G. "Exploration and Exploitation in Organization Learning." *Organizational Science* 2, no. 1 (February 1991): 71–87.
- Merton, Robert K. "The Unanticipated Consequences of Purposive Social Action." *American Sociological Review* 1, no. 6 (December 1936): 894–904.
- Mill, John Stuart. *On Liberty*. Indianapolis: Hackett Publishing, 1978.
- Miller, John H., and Scott E. Page. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton, NJ: Princeton University Press, 2007.
- Mitchell, Melanie. *Complexity: A Guided Tour*. New York: Oxford University Press, 2009.
- Muldoon, Ryan. *Social Contract Theory for a Diverse World: Beyond Tolerance*. New York: Routledge, 2016.
- Müller, Julian F. *Political Pluralism, Disagreement, and Justice: The Case for Polycentric Democracy*. New York: Routledge, 2019.
- North, Douglass C. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press, 1990.
- . *Understanding the Process of Economic Change*. Princeton, NJ: Princeton University Press, 2005.

- OECD. *Debate the Issues: Complexity and Policy Making*. Paris: OECD Publishing, 2017.
- Ostrom, Elinor. *Understanding Institutional Diversity*. Princeton, NJ: Princeton University Press, 2005.
- Page, Scott E. "The Imperative of Complexity." *Cosmos + Taxis* 5, no. 2 (2018), 4–12.
- Platteau, Jean-Philippe. *Institutions, Social Norms, and Economic Development*. New York: Routledge, 2000.
- Popper, Karl. *The Open Society and Its Enemies*. Princeton, NJ: Princeton University Press, 2013.
- Räikkä, Juha. "The Feasibility Condition in Political Theory." *Journal of Political Philosophy* 6, no. 1 (March 1998): 27–40.
- Rawls, John. *The Law of Peoples*. Cambridge, MA: Harvard University Press, 1999.
- . *Political Liberalism*. New York: Columbia University Press, 2005.
- . *A Theory of Justice*. Rev. ed. Cambridge, MA: Belknap Press, 1999.
- Riker, William R. *Liberalism against Populism: A Confrontation between the Theory of Democracy and the Theory of Social Choice*. San Francisco: W. H. Freeman and Co., 1982.
- Robeyns, Ingrid. "Ideal Theory in Theory and Practice." *Social Theory and Practice* 34, no. 3 (July 2008): 341–62.
- Rosenberg, Alex. "On the Very Idea of Ideal Theory in Political Philosophy." *Social Philosophy and Policy* 33, no. 1–2 (Fall 2016): 55–75.
- Schelling, Thomas. "Dynamic Models of Segregation." *The Journal of Mathematical Sociology* 1, no. 2 (1971): 143–86.
- Schmidtz, David. "Nonideal Theory: What It Is and What It Needs to Be." *Ethics* 121, no. 4 (July 2011): 772–96.
- . "A Realistic Political Ideal." *Social Philosophy and Policy* 33, nos. 1–2 (Fall 2016): 772–96.
- Sen, Amartya. *The Idea of Justice*. Cambridge, MA: Belknap Press, 2009.
- Shadish, William R., Thomas D. Cook, and Donald T. Campbell. *Experimental and Quasi-Experimental Design for Generalized Causal Inference*. Boston: Houghton Mifflin, 2002.
- Simmons, A. John. "Ideal and Nonideal Theory." *Philosophy and Public Affairs* 38, no. 1 (January 2010): 5–36.
- Simon, Herbert A. "The Architecture of Complexity." *Proceedings of the American Philosophical Society* 106, no. 6 (December 1962): 467–82.
- Smith, Adam. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: Methuen and Co., 1776.
- Smith, Peter. *Explaining Chaos*. Cambridge: Cambridge University Press, 1998.

- Stuntz, William J. "Self-Defeating Crimes." *Virginia Law Review* 86, no. 8 (November 2000): 1871–99.
- Swift, Adam. "The Value of Philosophy in Nonideal Circumstances." *Social Theory and Practice* 34, no. 3 (July 2008): 363–87.
- Tenner, Edward. *Why Things Bite Back: Technology and the Revenge of Unintended Consequences*. New York: Alfred A. Knopf, 1996.
- Tetlock, Philip E. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press, 2006.
- Valentini, Laura. "Ideal vs. Non-Ideal Theory: A Conceptual Map." *Philosophy Compass* 7, no. 9 (August 2012): 654–64.
- Wiens, David. "Against Ideal Guidance." *The Journal of Politics* 77, no. 2 (April 2015): 433–45.
- . "Political Ideals and the Feasibility Frontier." *Economics and Philosophy* 31, no. 3 (November 2015): 447–77.
- . "Prescribing Institutions without Ideal Theory." *The Journal of Political Philosophy* 20, no. 1 (March 2012): 45–70.
- Wilson, David S. "Two Meanings of Complex Adaptive Systems." In *Complexity and Evolution: Toward a New Synthesis for Economics*, edited by David S. Wilson and Alan Kirkman, 31–46. Cambridge, MA: MIT Press, 2016.

FREEDOM AND ACTUAL INTERFERENCE

Jonah Goldwater

DEBATES OVER THEORIES often concern how well rival theories explain paradigm cases. Debates over freedom are no different: because slavery is a paradigm of unfreedom, a theory's inability to adequately explain the slave's unfreedom can be used to reject the theory. This strategy is employed by Philip Pettit, who rejects the conception of freedom as noninterference—often called the negative or liberal conception—on the grounds that it cannot explain the unfreedom that slavery yields in at least one (type of) crucial case. That is the case of the so-called lucky slave: Pettit claims that if a slave had a kindly or well-meaning owner then that slave could be free from interference, rendering the slave free by the lights of the noninterference conception. Because this is absurd, and because the slave would remain dominated or controlled even if not interfered with, Pettit argues that freedom should be understood as the absence of domination—often called the republican conception—rather than the absence of interference.¹

Naturally, some have defended the noninterference conception.² Prominent among them are Ian Carter and Matthew Kramer, who have claimed that negative freedom—or at least the variant Carter and Kramer call “pure”—does indeed have the resources to explain the unfreedom of the lucky slave.³ For even if the slave is not actually interfered with, that the slave would be interfered with in certain hypothetical situations—such as the slave not engaging in the subservient behavior that results in the slaveowner's noninterference—suffices to show that the slave is unfree by the lights of the noninterference conception.

Carter and Kramer concede that this would render the slave's being inter-

- 1 See Pettit, *Just Freedom*, “The Instability of Freedom as Noninterference,” and *Republicanism*.
- 2 In addition to Carter (*A Measure of Freedom* and “How Are Power and Unfreedom Related?”) and Kramer (*The Quality of Freedom* and “Liberty and Domination”), see also Lang, “Invigilating Republican Liberty”; Wendt, “Slaves, Prisoners, and Republican Freedom”; de Bruin, “Liberal and Republican Freedom”; and Goodin, “Folie républicaine.”
- 3 Carter, *A Measure of Freedom* and “How Are Power and Unfreedom Related?”; Kramer, *The Quality of Freedom* and “Liberty and Domination.”

ferred with a matter of likelihood or probability, as well as a contingent empirical fact rather than a conceptual necessity.⁴ Is this an adequate explanation of unfreedom? Pettit thinks not. For according to Pettit there is a necessary or conceptual relation between slavery and unfreedom, not merely an empirical or contingent one, as well as a necessary or conceptual relation between democracy or republican government and (un)freedom. For on Pettit's view one either has the status and rights of a free person, which entails freedom from control or domination, or else one lacks that status, in which case one is subject to such control even if interference does not actually occur. So Carter and Kramer are also moved to deny that (un)freedom and forms of government are connected so intimately; on their view, just as the slave's unfreedom is contingent and empirical, so too is it only contingent and empirical that democracy promotes freedom more than, say, fascism.⁵

Though I suspect that Pettit is right here, the relation between freedom and forms of government is not the main concern of this paper (though I will return to it in the final section). Instead, I will argue that the agreement with Pettit that Carter and Kramer concede earlier in the dialectic—that a slave could be free from actual interference even if subject to possible or likely interference—is a mistake. My central claim is that the scope of actual interference is wider than has been recognized, a crucial implication of which is that there cannot be a slave with a noninterfering slavemaster any more than there can be a prisoner with a noninterfering jail cell. So against Pettit I argue that the noninterference conception of freedom does indeed have the resources to explain the unfreedom induced by slavery. In keeping with Pettit, however, my position satisfies the stronger demand that Carter and Kramer attenuate: that a slave is necessarily rather than contingently unfree—the reason, I claim, is that being a slave *entails* being interfered with. This result obviates the debate over whether a high probability of possible interference becoming actual is sufficient to explain the slave's unfreedom.

This paper has a second goal as well. In addition to arguing that domination and unfreedom can occur without actual interference, Pettit also claims that one can be free while interfered with if one is not dominated. In particular, if laws are produced via the legitimate procedures of a well-ordered democracy or republic, then even if one is interfered with by such laws, one is not dominated if

- 4 As Carter puts it, "people who are subject to [domination] can be seen as less free in the negative sense even if they do not actually suffer interference, because the probability of their suffering constraints is always greater (*ceteris paribus*, as a matter of empirical fact) than it would be if they were not subject" to that domination ("Positive and Negative Liberty," sec. 3.2).
- 5 For an argument that this implication is especially problematic, see Harbour, "Non-Domination and Pure Negative Liberty."

those laws track the “avowed interests” of the republic’s citizens.⁶ In fact, Pettit even holds that democratic laws may promote rather than hinder freedom. So a further reason to adopt the non-domination view of freedom, Pettit argues, is that it, unlike the noninterference view, can explain how laws and regulations can be liberating rather than oppressing. Though I agree that laws and regulations can indeed be liberating, I also reject Pettit’s claim that the noninterference conception of freedom is unable to explain this. The core reason is that even if a law interferes with some action, it can nonetheless protect against even greater interference, all things considered. That is, just as according to the utilitarian an action can cause some pain but a greater amount of pleasure overall, so too can a law interfere to some degree but still be freedom enhancing on balance or *ultima facie*. Contrary to what many believe, then, I show that laws and regulations can enhance negative freedom rather than simply impede it.

The paper is structured as follows. In section 1, I defend my thesis regarding the wide scope of actual interference. In section 2, I apply that thesis to the arguments for freedom as non-domination. In section 3, I show how laws and regulations may enhance rather than impede (negative) freedom, and in section 4, I apply this result to questions about forms of government. The overall result: Pettit’s charge that the noninterference conception cannot meet its explanatory burden is unfounded.

1. THEORIES, AUXILIARY ASSUMPTIONS, AND ACTUAL INTERFERENCE

Above I noted that debates over theories often turn on the explanation of paradigm cases. I also indicated that my central thesis concerns the scope of interference. Both can be illustrated by what I take to be a paradigm of an unfree person—namely, a person in prison.⁷ How would the noninterference conception explain the prisoner’s unfreedom? The answer appears simple: the prisoner is

6 See Beckman and Rosenberg, “Freedom as Non-Domination and Democratic Inclusion,” for a recent discussion of democracy, citizenship, and non-domination.

7 Interestingly, Pettit actually denies that imprisonment is paradigmatic of unfreedom; for discussion, see Wendt, “Slaves, Prisoners, and Republican Freedom.” Though space precludes an in-depth discussion, it is worth noting that Pettit is pushed this way as an implication of his views, including his criticisms of negative freedom. Though denying the paradigmatic nature of the prisoner’s unfreedom might ultimately be justifiable via theoretical considerations or an appeal to reflective equilibrium, say, if one can avoid denying a paradigm case I assume one should. One reason is general and meta-theoretical: insofar as explaining paradigm cases is an important means of comparing theories, denying a paradigm threatens to undercut the prospects of theory-neutral assessment. A second reason, particular to this case, is that insofar as my arguments (throughout the paper) undercut the reasons for abandoning negative freedom while maintaining the paradigmatically unfree character of

interfered with. But when, and how often, is the prisoner interfered with? Is the prisoner unfree only when he or she pushes or struggles against the jail walls? The answer seems obvious; whether the prisoner struggles or not, the prison walls interfere with the range of choices the prisoner might otherwise make. So it would seem the prisoner is interfered with or constrained, and so not free, for the entirety of the prison sentence.

I assume that if a theory misjudges or cannot explain a paradigm case the theory should be rejected (all else equal). Famously, the possibility of a contented slave is thought to refute the theory of freedom as the ability to do what one wants; rather than thinking that a slave who learns to want only what he can have thereby liberates himself, the idea is that freedom should not be indexed to what one happens to want.⁸ So in a similar vein suppose one held that the jail cell interferes with the prisoner only if the prisoner actively struggles against the bars. Then the key to liberation would be to sit still; on this view, even a person confined to a cell just barely larger than that person's body would be counted as free as long as that person did not move a muscle. So according to freedom as noninterference conjoined with the view that interference occurs only when one physically butts up against obstacles or constraints, it would seem that the less one moves, the freer one is.⁹

The absurdity of this verdict might tempt one to reject freedom as noninterference, just as the slave learning not to want what he cannot have is a reason to reject the theory of freedom as the ability to do what one wants. But such a criticism aims at the wrong target. To see this, consider the general distinction between a theory and an auxiliary assumption familiar from the philosophy of science. Whereas a theory proper consists only of central or core claims, auxiliary assumptions are supplementary claims used in conjunction with the core theory in order to derive specific predictions or verdicts. Applying the distinction here suggests that one should distinguish the core theoretical claim that freedom consists in noninterference with the auxiliary assumption that actual interference occurs only if one struggles against an obstacle such as a jail wall. Conjoined, these yield the prediction or verdict that a person who learns to sit still in jail thereby liberates himself. That this prediction or verdict is false need

imprisonment, there is no need to adopt the counterintuitive position that Pettit adopts here.

- 8 The problem was first raised by critics of Berlin's famous *Two Concepts of Liberty*, such as Benn and Weinstein, "Being Free to Act, and Being a Free Man." See Flikschuh, *Freedom*, ch. 1, sec. 2.3, for an overview.
- 9 The thought calls to mind a line often attributed to activist and theorist Rosa Luxemburg: "those who do not move do not notice their chains."

not impugn the theory, however, but should instead impugn the auxiliary assumption. For the same theory conjoined with a more plausible auxiliary assumption—that a jail wall interferes with someone even when not butting into it, for example—yields the correct prediction or verdict, namely, that even a person who learns to sit still in prison remains unfree.

With the distinction between theory and auxiliary assumption in mind consider another paradigm case. I take it as a truism or a paradigm of unfreedom that one is unfree to do x if it is illegal to do x .¹⁰ Yet suppose instead that one is rendered unfree to perform an illegal x only when one is arrested for doing x . Then the path to liberation with respect to an illegal x would simply be to refrain from doing it, thereby avoiding arrest; more succinctly, one would be free to do an illegal x as long as one did not do it. But this verdict is just as absurd as the idea that the contented slave liberates himself by learning to not want to do what he is barred from doing, or that a prisoner can liberate himself by sitting still and not touching the bars. Instead, the more plausible thought is that one is unfree to do x whenever or wherever it is illegal, not only when one is caught, just as one is unfree whenever one is in prison, not only when one struggles against the bars.

So, to briefly summarize. One might reject freedom as noninterference on the grounds that it implies that a prisoner could liberate himself by sitting still, or that a citizen could liberate himself by not performing an illegal action. While I agree that a theory with these implications should be rejected, I argue that it is not the noninterference theory that has these implications, but it is this theory conjoined with a (perhaps implicit) auxiliary assumption that restricts interference to moments of struggle or arrest; call this the “narrow-scope” reading of interference. Instead, a different auxiliary assumption that recognizes a wider scope of actual interference—one that counts jail walls and laws as interfering even when not butting into them or being arrested in light of them—delivers the correct verdicts, and the correct explanations. Thus, the truisms and cases discussed so far supply an argument for the “wide-scope” conception of actual interference: whereas the narrow-scope conception absurdly implies that one can liberate oneself by sitting still (the “passive prisoner” case), or by refraining from illegal actions (the “upstanding citizen” case), the wide-scope conception delivers the correct verdicts in cases such as these.¹¹ Therefore “interference” should be understood in the wide not narrow sense.

10 This is of course compatible with being free to do x in the sense of “free will”; having the free will to choose to break the law does not entail that one is free to break the law in the social or political sense, and it is only with the latter sense of “freedom” that I am concerned here.

11 My thanks to an anonymous referee for the *Journal of Ethics and Social Philosophy* for the “passive prisoner” moniker.

What objections might be raised here? One objection concedes that jail walls interfere in the wide sense, but holds that laws do not. I will return to this objection shortly. Instead, first consider objections to even jail walls interfering in the wide sense, and for now assume that jail walls and laws are relevantly similar—an assumption that will soon be discharged.

An initial objection to the wide-scope conception appeals to the notion of a disposition: perhaps one might think jail walls or laws, even if actual, only have a disposition to interfere without actually interfering, and only actually interfere when the dispositions are manifested (by physical contact or enforcement, say). But this objection fails, as it conflates the disposition/manifestation distinction with the possible/actual distinction. To illustrate: glass is fragile, which implies that it has the disposition to break easily. But this does not imply that if it is not breaking the glass is only *possibly* rather than *actually* fragile. The reason is that dispositions are actual even when not manifesting. So even if the jail wall's impenetrability—its disposition to repel solid matter—is not manifesting, the jail wall actually has the disposition. And, of course, it is for that very reason that a prisoner cannot actually walk through the walls even if he tries. So the natural conclusion is that the jail's actual impenetrability, even without manifesting, actually interferes with what one can do, and therefore with what one is free to do.¹²

A related objection appeals to a hypothetical or subjunctive construal: in particular, one might think that saying the bars would interfere if one were to struggle against them entails or is tantamount to the bars only possibly rather than actually interfering. In fact, note that this line of reasoning is what motivated Carter and Kramer's concession to Pettit, as described at the outset: although they defend a negative or noninterference conception of freedom against Pettit, Carter and Kramer nonetheless follow Pettit in thinking that hypothetical or subjunctive interference is tantamount to possible rather than actual interference (even if, *contra* Pettit, they go on to hold that the high probability of possible interference becoming actual is sufficient for attributing unfreedom).

But this concession is unnecessary. For such hypothetical or subjunctive claims are derivative rather than basic or fundamental. After all, the reason one would be prevented from leaving the cell if one tried is simply that the cell actually exists, and actually has the dispositions it has. Put another way, it is the cell's actual existence and nature that is the ground or truthmaker for claims about what would happen were one to struggle against it. Therefore it is the actual exist-

12 It seems to me that Lang ("Invigilating Republican Liberty," 288) suggests something roughly similar.

tence and nature of the actual cell that interferes with what one can actually do.¹³ It should also be emphasized that restricting the range of possible choices does not imply that the jail cell only possibly rather than actually interferes. For it is the choices here that are possible, not the interference. Put another way, there is a difference between a possible interference with an actual choice and an actual interference with a possible choice. And not only is it the choices that are merely possible, not the interference, it is the jail's actual interference that renders certain choices non-actual.

Two more reasons are worth identifying as a source of reluctance here if any still are felt. One is a tendency, if only on occasion, to think of the possible as the counterfactual, and the counterfactual as the possible. Yet hypothetical or subjunctive claims need not be *counter* to fact as opposed to consistent with it, and the interference of jail walls and laws need not concern past-facing counterfactuals (what could have gone differently?). Jail walls and laws can instead invoke or imply claims about the future, such as what would actually happen if one were to run afoul of them.¹⁴ The second reason is a perhaps implicit tendency to treat "possible" as mutually exclusive with "actual," and "merely possible" as akin to "possible." But actual and possible are not mutually exclusive; trivially, anything actual is possible, and future possibilities may be consistent with actuality, as just noted. There is also good reason to distinguish "possible" from "merely possible." For example, it seems inappropriate to describe "it is possible it will rain tomorrow" as invoking a "mere" possibility given that this expression may invoke nothing but ignorance of actual meteorological conditions and laws of nature. The possibility that unicorns or leprechauns could exist, by contrast, does seem fairly described as a "mere possibility." But if the phrases work this way (roughly speaking) then using "merely possible" regarding the prospect of manifesting interference may obscure the way future events need not contrast with actuality as mythical creatures do.

To avoid the charge of diagnosing a straw man, consider an important example tying together several of the issues just discussed. Like Carter and Kramer, Goodin and Jackson defend the move toward understanding unfreedom in

13 To forestall the objection that only human agency can take away freedom, such that jail-cell walls cannot, it need only be kept in mind that humans design and maintain the jail for just this purpose.

14 If one assumes such claims invoke possibility due to being analyzed via possible worlds, it may be worth emphasizing that possible-world talk can be quite misleading; assuming modal realism is false, all modal facts are ultimately grounded in the actual world—as there is no other world, literally speaking, for them to be grounded in. So the invocation of possibility need not invoke non-actuality in the sense at issue.

terms of probable or likely interference rather than actual interference.¹⁵ Their paper “Freedom from Fear” begins as follows:

We think our house would look nice painted white. A city ordinance prohibits that. What is it precisely that impinges on our freedom to paint our house white? When should we rationally fear the city’s interference? One answer might be that it is only the city’s *actual interference* that impinges on our freedom. But the effect on our freedom happened well before then. Should we rationally fear only when the constable knocks at the door, handcuffs in hand? The threat was real, and we should have feared it, long before *the actual knock* at the door. Is it the *mere possibility* of interference that impinges on our freedom? Should we rationally fear whenever there is a chance that a constable might possibly appear? That seems premature. Before taking fright, we ought rationally ascertain the likelihood of that possibility, which might turn out to be very remote.¹⁶

First, while Goodin and Jackson are right that a law prohibiting white paint affects one’s freedom to paint something white even before the policeman knocks, they assume without argument that the “actual knock” is equivalent to or coextensive with “actual interference.” Second, their not distinguishing theoretical claims from auxiliary assumptions affects their framing of the issue, as they take the problem they identify as a problem for the theory of freedom in terms of actual (non)interference, rather than a reason to keep the theory but change the auxiliary assumption regarding what counts as actual interference. Third, Goodin and Jackson seem to conflate “possible” with “merely possible” in the manner described above. For the prospect of police enforcing a law is a very real possibility in a way that unicorns and leprechauns are not. This is no small point, as this framing is what motivates Goodin and Jackson’s “probabilist” account: against the “actualist” who thinks freedom is “the absence of any actual external impediments to action,” and against the “possibilist” who thinks freedom is “the absence of any possible external impediments to action,” Goodin and Jackson advocate “probabilism,” according to which freedom is “the absence of any probable external impediments to action.”¹⁷ But even assuming probabilism is the best of these three *given Goodin and Jackson’s premises*, these premises embody a too-narrow construal of actual interference (or so I am arguing). So rather than argue for or against probabilism so construed, my arguments, if successful,

15 Goodin and Jackson, “Freedom from Fear.”

16 Goodin and Jackson, “Freedom from Fear,” 249 (emphasis added).

17 Goodin and Jackson, “Freedom from Fear,” 251.

would obviate the need to choose among the options in Goodin and Jackson's trichotomy.

With an eye toward returning to the non-domination view of freedom in the next section, consider another way of making the case for the wide-scope construal of actual interference. Perhaps some find it natural to think of interference as an event, in particular a momentary event, perhaps akin to the manifestation of a disposition. Yet for those who think unfreedom can occur due to domination without interference, domination seems to be construed as an enduring state, a systemic state of affairs, or a "structural relationship," as Costa puts it.¹⁸ For example, in his defense of freedom as non-domination, Skinner contrasts "interference or even any threat of it" with "the predicament" of those who live "in subjection to the will of others," which, by its very nature, "has the effect of placing limits" on liberty.¹⁹ It is not clear that Skinner's contrast between a "predicament" and what appears to be more short-lived moments of interference is justified, however. Instead, just as domination is thought to exist as an enduring predicament to which one is subject even when not manifesting in specific acts of domination, so too should one think of being subject to another's will as an enduring state of interference, one that exists (perhaps as a disposition) even when not manifesting in specific acts of interference. So, in brief, I suggest that one can, and should, think of interference just as one thinks of domination—as an ongoing or enduring state of affairs that constrains what people can choose to do for the entirety of its duration.

Lastly, return to the objection mentioned but deferred earlier: that even if I am right about the wide scope of a jail wall's interference, one might think laws do not interfere in the same way. For not all laws are (actually) enforced. For instance, some laws might purport to apply beyond their authority; consider a small town declaring something illegal at the federal level, or a single country declaring something an international crime. Other laws might go unenforced even within their jurisdiction—perhaps due to a lack of resources, or because a law still on the books seems antiquated by contemporary mores. Yet in either case one might think my view—unlike other theories of freedom, such as the non-domination conception or Goodin and Jackson's probabilist view—would wrongly treat people as being interfered with even by unenforced laws, and so rendered unfree by them.²⁰

But this is not the case. The core reason is that unenforced laws are analogous

18 Costa, "Neo-Republicanism, Freedom as Non-Domination, and Citizen Virtue," 406.

19 Skinner, "A Third Concept of Liberty," 262–63.

20 My thanks to an anonymous referee for the *Journal of Ethics and Social Philosophy* for this objection.

to physically ineffective jail walls. Suppose it turns out that one could tunnel out of jail, or wiggle in between the surprisingly loose bars in one's cell. This would not show that walls do not interfere *in general* or in the typical case, as argued above. It would only show that these particular walls are not as effective at interfering as they are intended to be. Consider this in light of the other rebuttals given above. Although walls are typically impenetrable—i.e., they typically have the disposition or ability to repel solid matter—a particular wall through which one could escape would lack this disposition, or else have it to a degree insufficient to repel the escapee (even while repelling other solid matter). Generally speaking, interference lends itself to being understood as a physical force, or in physical terms. So the relative strength or magnitude of this force—as well as the various opposing forces—is often relevant. That Superman could bend the bars of a cell, or a prisoner equipped with dynamite could explode his way out, does not imply that walls do not generally interfere, nor that someone in normal circumstances is not interfered with and so is unfree when imprisoned by those same walls. Such cases only show that the degree or extent of interference is not infinite or necessarily insurmountable. But then someone could be free after all if surrounded by walls too weak to imprison.²¹

All of this applies to laws, *mutatis mutandis*. An enforced law is analogous to a functioning wall, and an unenforced law is analogous to a wall through which a prisoner could escape. In both cases the general or typical state of affairs is one of (physically) overwhelming interference, sufficient for unfreedom. Yet this is perfectly compatible with there being physically faulty particular instances, such as certain walls weak enough to tunnel through or certain laws that go unenforced, rendering a particular person free in those particular circumstances, with respect to those walls or laws.

Consider the rebuttal one last way. Note that my arguments for the wide-scope interference of jail walls would equally apply were the walls replaced with human guards surrounding a prisoner on all sides. For such a formation would have the same effect as a wall, and possess relevantly similar dispositions. Yet even if the guards opted to let the prisoner go (something a wall could not do), this would be no different, with respect to freedom, than the guards opting to not ensure the jail cell was physically escape proof. So not enforcing a law would be similar: what normally yields a sufficient degree of interference for unfreedom would not in such cases.

21 Note that this (correctly) implies that whether someone is unfree can depend on the physical capacities of that person. For instance, Superman in a jail cell would be just as free as if he were in a hotel room, even if an ordinary man would find these rooms to relate to freedom quite differently.

So to briefly summarize the results of this section. I have argued for the wide-scope conception of actual interference by appealing to the superior verdicts it generates; unlike the narrow-scope construal that wrongly implies that the passive prisoner is liberated by sitting still, for example, the wide-scope construal correctly implies that the prisoner is interfered with even if passive. Furthermore, because the objections to the wide-scope conception fail, this provides further argumentative support for the wide-scope conception going forward.

2. SLAVERY, DOMINATION, AND ACTUAL INTERFERENCE

It was mentioned at the outset that Pettit rejects freedom as noninterference on the grounds that one can be unfree without being interfered with if dominated, and that one can be interfered with yet free if not dominated. My focus in this section is with the first claim. In the next section I will turn to the second.

Why does Pettit think that one can be dominated without being interfered with? As described at the outset, the answer appeals to what is often called the case of the lucky slave. In short, the core idea is that the slave who is lucky to have a kind owner might well be free to make choices without the owner's interference. Because such a slave is not actually interfered with, the idea goes, the lucky slave would be deemed free by the theory of freedom as noninterference. But, of course, a slave is not free. So freedom cannot consist in noninterference. Instead, because the slave remains dominated even if not interfered with, according to Pettit freedom should be understood as the absence of domination.

For reasons that will be clear shortly, it is worth quoting Pettit several times from his book *Republicanism*:

It is possible to have domination without interference and interference without domination. I may be dominated by another—for example, to go to the extreme case—I may be the slave of another—without actually being interfered with in *any* of my choices. It may just happen that my master is of a kindly and non-interfering disposition. Or it may just happen that I am cunning or fawning enough to be able to get away with doing *whatever* I like.²²

There may be a loss of liberty without *any* actual interference: there may be enslavement and domination without interference, as in the scenario of the non-interfering master.²³

22 Pettit, *Republicanism*, 22 (emphasis added).

23 Pettit, *Republicanism*, 31 (emphasis added).

Slavery is essentially characterized by domination, not by actual interference: even if the slave's master proves to be *entirely* benign and permissive, he or she continues to dominate the slave.²⁴

The enjoyment of dominating power over another is consistent with *never* actually interfering.²⁵

You can be dominated by someone, as in the case of the lucky or cunning slave, without actually suffering interference at their hands.²⁶

One may note that I have emphasized Pettit's frequent use of unrestricted or general words such as "never," "any," "entirely," and the like; rather than suggesting a slave might be lucky with respect to noninterference in a certain specific range of choices, Pettit consistently claims that it is possible for a slave to do "whatever" he likes.

Nonetheless, when others describe the lucky-slave case their language is more hedged or qualified. For example, List and Valentini present the case as follows:

Critics [of freedom as noninterference] have argued that [the] focus on actual constraints . . . has problematic implications. Consider the often-cited case of a slave with a *noninterfering master*. In this hypothetical scenario, the master could in principle interfere with the slave's actions . . . but it so happens that the master refrains from interfering, and *many* actions are actually open to the slave. On the liberal conception, the slave would count as free to perform *these actions*—a conclusion that many find unsatisfactory in light of the paradigmatically *unfree status* associated with slavery.²⁷

Lang presents the case similarly:

Think here of a slave and a kindly master. The master may be kindly disposed towards his slave, and offer *little* actual interference with her activities. Nonetheless, the master could, with impunity, interfere with every detail of his slave's life, if he were so disposed. For republicans, the slave is unfree even if she is *not actually interfered* with.²⁸

24 Pettit, *Republicanism*, 32 (emphasis added).

25 Pettit, *Republicanism*, 65 (emphasis added).

26 Pettit, *Republicanism*, 80.

27 List and Valentini, "Freedom as Independence," 1052 (emphasis added).

28 Lang, "Invigilating Republican Liberty," 275 (emphasis added).

And for a third example, consider Carter:

Freedom is not simply a matter of non-interference, for a slave may enjoy *a great deal of* non-interference at the whim of her master. What makes her unfree is her status, such that she is permanently liable to interference of any kind. Even if the slave *enjoys non-interference*, she is, as Pettit puts it, “dominated,” because she is permanently subject to the arbitrary power of her owner.²⁹

So, whereas Pettit had described the slave as being entirely free from interference such that the slave could do “whatever” he liked, each of the three passages here is more qualified, at least at first: List and Valentini present the slave as being free to perform “many” or “these” actions rather than any or all; Lang starts by suggesting that the slave suffers “little” noninterference but not none; while Carter initially describes the slave as enjoying “a great deal of noninterference” but not noninterference *tout court*, or unqualified.

Although these restrictions fall short of Pettit’s own description of complete or total noninterference, it is instructive to consider each version. Suppose first that the noninterference is restricted rather than total. But then there is not even a *prima facie* reason to think that the noninterference conception wrongly implies the slave should be considered free (*pace* the inference drawn in the cited passages after describing limited interference). For even if the slave is not interfered with regarding some or certain actions, if the slave is still interfered with regarding other actions then the slave will still count as unfree by the lights of the noninterference conception. (This is especially so if the permitted actions are mundane or quotidian, whereas the blocked actions are more fundamental or existentially important, such as where to live or raise one’s children.) Consequently, even if Pettit is right to focus “on what it is to enjoy freedom as a person or citizen” as opposed to “freedom in a particular choice or type of choice,” the conception of freedom as non-domination gets no traction in the restricted variant of the case.³⁰

So for the lucky-slave case to do its intended work one must construe it as Pettit initially formulates it, with the noninterference considered total. But the crucial problem here is that it is hard to see how a case of total or complete non-interference is even possible. Consider: while it is certainly plausible to imagine a slavemaster who is not an inveterate micromanager, explicitly ordering what the slave is to do at every moment and with respect to every decision, it is much harder to imagine a globally or universally noninterfering slavemaster, as Pettit

29 Carter, “Positive and Negative Liberty,” sec. 3.2 (emphasis added).

30 Pettit, *Just Freedom*, 29.

suggests might be the case. Suppose the slave wants to pursue higher education or practice the religion of his choice. Or suppose the slave wants to speak out publicly against the evils of slavery, refuse to let his children be sold at auction, or not be a slave anymore. Is one to imagine a master, even a benevolent or kindly one, allowing his slaves to seek employment elsewhere, or to choose not to be slaves? If the answer is yes then the slave is no longer a slave, and so *a fortiori* no longer a lucky slave (who remains dominated *qua* slave). But if the answer is no then the slave is interfered with after all. So it would seem that the only way for complete or total noninterference to be possible is to imagine that the slave does not or would not *want* to choose or ask for such things—for example, that the slave does not want to not be a slave, or does not want to be educated, or does not care if his children are sold at auction. That is, one must imagine the slave only wanting to do the things that the master will let the slave do, rather than imagine the slave wanting to do the things that any free (non-enslaved) person would likely want to do—or at least be free to do. What this suggests, then, is that Pettit is running together the case of the *lucky* slave with the case of the *contented* slave.

As described earlier, the case of the contented slave is intended as a *reductio* of the theory of freedom as the ability to do what one wants: if freedom is the ability to do what one wants, the idea goes, then if one learned to want only what one could get—as in the case of the slave who learns to want only what the owner will allow him—then all of one's desires could be satisfied, and one would be free. But one cannot liberate oneself simply by adapting one's preferences or desires—such an idea is absurd—so therefore freedom cannot simply be the ability to do what one wants. Yet the case of the lucky slave is supposed to be different. For the case of the lucky slave involves the dispositions and preferences of the master, not the slave; it is the master who is benevolent and so noninterfering, not the slave who adapts his preferences. Yet Pettit seems to run these together: only if the slave is content wanting what the master will not interfere with can the slave be said to be free from interference. But then the lucky slave case cannot do the argumentative work it is intended to do. For surely part of the lesson of the contented-slave case is that whether someone is free is not simply a function of their ability to satisfy the desires they actually have. But then this should carry over to the lucky-slave case: being interfered with is not simply a function of the inability to act on the desires one actually has. Instead, actual interference occurs when a range of possible choices—choices one would otherwise make—are blocked, regardless of whether one actually wants them, or whether one has learned to not want them by adaptive preference formation.

Recall that earlier I argued that if it is illegal to do x one is thereby unfree

to do x , not only when one is arrested, and not based on whether one wants to do x . For the law interferes regardless of one's preferences. The same ideas apply here. Assuming that the slavery in question is a legal institution, the slave is legally barred from doing many of the things that one is free to do in a society that outlaws slavery. In a slave-holding society it may be illegal for the slave to publicly denounce slavery, receive an education, or hide his children from the auction block, say. (And sadly, such laws are typically if not brutally enforced.) So contrary to Pettit's assumptions, the individual master's dispositions are irrelevant. For even if the individual master lets the slave do something illegal the slave is still not free to do that something, any more than a Mafia don giving me the go-ahead to murder someone makes me free to murder. For if laws actually interfere then one is unfree to murder as long as there is a law against it, no matter what a crime boss says. So if laws actually interfere then the slave is interfered with at every moment that slavery is legal and that person is legally a slave—no matter what the individual master says the slave can or cannot do. So the master allowing the slave some leeway—what Pettit later calls “free rein” as opposed to the freedom of being “unharnessed”—is simply not enough to show that it is possible for a slave to never be interfered with, rather than, on certain occasions, not be interfered with in respect to a certain small range of choices that are not legally proscribed by the institution of slavery.³¹

So in effect my claim here is that the lucky-slave case—in which a slave is never interfered with yet remains a slave—is an impossibility. Put another way, just as there can be no such thing as noninterfering prison walls—if you are in the cell the walls interfere—there can be no such thing as noninterfering legalized slavery—if you are a slave the law interferes.³² So Pettit's lucky-slave argument falls to a presupposition failure, and so fails to motivate freedom as non-domination against freedom as noninterference.

31 Pettit, *Just Freedom*, 3.

32 Given the earlier-discussed complication of enforceability akin to the (im)penetrability of a wall, the phrasing here should be construed as shorthand for a more complex formulation invoking something akin to “physically functioning walls in normal circumstances,” or “functioning laws in normal circumstances.” Lest one think this weakens or undercuts my point here, however, compare the case with Pettit's own view. Suppose it is written into law that slaves or other groups of people are to be subjugated or dominated as second-class citizens. Yet further suppose that such laws are not enforced, and everyone enjoys a *de facto* equality. Are such people still dominated and so unfree? If so, then even unenforced domination (that in a sense only exists on paper) suffices for unfreedom. But then the same should hold for my account: unenforced interference suffices for unfreedom as well. Or else suppose that such domination would be merely nominal and not amount to unfreedom. But then that too can be applied to my account; for as argued earlier, unenforced laws need not take away freedom any more than walls one can bypass need confine.

Does this mean the slave is not dominated, however? Of course not: especially given my arguments for the wide scope of actual interference, rather than treat domination as *contrasting* with interference, as Pettit does, one should instead think of domination as a particularly intense or insidious form of interference; the institution of slavery is such a ubiquitous actual interference in the lives of slaves that it is eminently reasonable to describe slaves not just as governed by the laws of slavery, but dominated by them.

Note too that the same arguments apply, *mutatis mutandis*, to the case of the dominated wife under the laws of “coverture,” which gave a husband broad legal authority over his wife. This is worth emphasizing because instead of the lucky slave, Pettit invokes a “lucky wife,” as it were, to motivate freedom as non-domination in the prologue to his 2014 book, *Just Freedom*. In particular, Pettit invokes the fictional characters Torvald and Nora from Henrik Ibsen’s play *A Doll’s House*. As Pettit puts it,

under nineteenth-century law Torvald has enormous power over how his wife can act, but he dotes on her and denies her nothing—nothing, at least, within the accepted parameters of life as a banker’s wife.... When it comes to the ordinary doings of everyday life, then, Nora has carte blanche. She has all the latitude that a woman in late nineteenth-century Europe could have wished for.³³

In light of my arguments above, however, one may notice that Pettit’s qualifications here make all the difference; Pettit’s invocation of “accepted parameters,” and the restriction to “the ordinary doings of everyday life,” clearly invoke a circumscribed range of action beyond which Nora cannot be said to freely choose.³⁴ Yet despite acknowledging these restrictions—restrictions that, I argue, constitute actual interference with Nora’s range of choices—Pettit nonetheless claims that by the lights of freedom as noninterference Nora is free, an incorrect judgment that motivates adopting the non-domination conception of

33 Pettit, *Just Freedom*, xiv.

34 Note hints of the above-discussed conflation of the lucky slave with the contented slave; to say Nora has “all the latitude” a woman in *that time and place* could have wished for, when, quite clearly, a twenty-first-century American woman would wish for so much more, just goes to show that at least some of Nora’s preferences must have been adapted to her limited circumstances—i.e., the area of action not actually proscribed by law. After all, what if Nora wishes to own her own property? Or seek a no-fault divorce? If this is beyond the pale for a nineteenth-century woman, that only goes to show that coverture laws constitute an actual interference in the range of women’s choices, such that it is hardly unsurprising for a woman in such a scenario to not wish for more—just as a slave resigned to slavery may no longer dare to dream of freedom.

freedom instead.³⁵ But as I have argued, this is simply unnecessary. Legalized coverture, like legalized slavery, amounts to a constant (and yes, dominating) actual interference in the lives of women, and slaves. So freedom as noninterference is perfectly well equipped to explain why women legally bound to their husbands are actually interfered with and so unfree, even if those husbands are indulgent or nice.³⁶

Lastly, recall that at the outset I noted that Carter and Kramer concede that a slave might be free from actual interference, such that the debate between freedom as non-domination as opposed to noninterference should turn on the question of whether the (high) probability of possible interference becoming actual is sufficient to explain the slave's unfreedom. Simply by recognizing the wide scope of actual interference, however, one can avoid this debate that starts from the viability of the lucky-slave thought experiment. In addition, recall also that Pettit criticizes Carter and Kramer for attenuating the link between slavery

35 Pettit, *Just Freedom*, xiv.

36 What, though, if the domination occurs outside the law? Perhaps even a wife with equal legal status may be dominated and so unfree, one might think. Similarly, perhaps one might think that an employee may be dominated by an employer—say, via implied threats or requests that cannot be refused—and so unfree as a result, all without actually being interfered with. My first response is meta-theoretical. Because whether such scenarios involve unfreedom in the first place is controversial, it cannot be assumed as a fault or benefit that a theory does or does not classify them as free; famously, the Marxist and capitalist disagree on whether poverty or market forces interfere with freedom, but, for that very reason, one cannot simply reject the capitalist theory because it fails to account for the datum that poverty renders one unfree, say. Second, but relatedly, a dispute over whether some *case* involves unfreedom need not be a dispute over *conceptions* of freedom: to use the same example, both Marxists and capitalists may assume freedom is noninterference but disagree on whether private property boundaries count as a relevant sort of interference (see Cohen, “Capitalism, Freedom, and the Proletariat”; Waldron, “Homelessness and the Issue of Freedom”). Instead, and as I have explained throughout the paper, the locus of the dispute may involve auxiliary assumptions about what counts as an actual interference, rather than the theory of freedom (as something other than actual interference). Third, and with that said, my account is neutral, or, at least, could be used by either side, to account for the freedom or its absence in the case of the legally equal but dominated wife, or in the case of the legally free but dominated employee. First, suppose that one thinks that the wife and employee are unfree because dominated. Then on my account whatever factors constitute that enduring state of domination, even apart from instances or acts of domination, can also be understood as ongoing or enduring interferences, just as I have argued that laws constitute enduring or ongoing interferences even when one is not arrested for breaking them. Alternately, suppose the wife and employee are free (perhaps because they are physically or legally able to leave without significant harm). Then, the idea goes, the obstacles here are not strong enough to prevent one from leaving, such that they do not adequately interfere with one's range of choices to count as restricting freedom—akin, more or less, to being free to refuse a coercive offer because the threat comes from a person too weak to carry it through.

and unfreedom: for Pettit that a slave is unfree is a necessary or conceptual truth, whereas for Carter and Kramer that a slave is unfree is contingent and empirical, thereby leaving open the genuine empirical possibility that a slave really could get lucky. My arguments show that this concession is unnecessary. Just as being in a jail cell *entails* unfreedom due to the jail's actual interference with one's range of choices, so too does being a slave entail unfreedom due to slavery's actual interference with one's range of choices. The constraints or obstacles that are jails and slavery are not merely empirically correlated with unfreedom.³⁷

3. INTERFERENCE, REGULATIONS, AND OVERALL FREEDOM

It was indicated at the outset that Pettit not only claims that there can be unfreedom if there is domination without interference, but also that there can be freedom with interference if there is no domination. I now turn to the latter claim.

To show how there can be freedom with interference if there is no domination, Pettit appeals to the example of traffic laws: while traffic laws might interfere with one's actions on the road, such laws do not take away freedom, Pettit claims, because they are not dominating. According to Pettit, more generally:

Government inevitably involves interference in the lives of citizens, whether via legislation, punishment, or taxation. Our [republican] ideal suggests that this interference need not be dominating, however—and need not be inherently inimical to freedom—so long as the people affected by the interference share equally in controlling the form it takes. Let state interference be guided equally by the citizenry and it will not reflect an alien power or will in their lives.³⁸

Because such legitimately enacted democratic laws do not amount to an alien will, such laws do not subject one to an alien will, and so one remains free under such laws. But Pettit also goes further. Not only is one not dominated or rendered unfree by democratically enacted laws, such laws might actually enhance one's freedom; for example, and as Pettit puts it, “the rules introduced under a

37 Moreover, this rebuts Carter's claim that only an empirical and contingent relation would not trivialize the opposition of freedom and slavery (“How Are Power and Unfreedom Related?” 81). For my claim that a prisoner is unfree is not a trivial tautology, but rather an implication of the empirical fact that a prisoner exists in a physical space that physically proscribes the range of choices the prisoner can make.

38 Pettit, *Just Freedom*, xx.

property or transport system, far from restricting freedom, may actually facilitate it.”³⁹

So Pettit concludes that not only does interference not suffice for unfreedom, but the interference of, for example, traffic rules might actually enhance freedom. Especially because I have argued that laws amount to an ongoing interference, however, one might think Pettit’s view has an advantage here. Expressed another way, it would appear that my account is susceptible to Pettit’s criticism: if a law amounts to an interference, as I contend, then even a legitimate, well-meaning, and democratically enacted just law amounts to a freedom-restricting interference, a thought which might push one toward a right-wing libertarian view that sees freedom as being maximized in the absence of laws. Yet, like Pettit, many on the left might think of (certain) laws as freedom enhancing, especially regulatory laws that protect citizens or consumers against excessive corporate power. It might then be thought an advantage of Pettit’s account that he has the resources to argue that “regulation is not oppression,” and, more strongly, that “we establish most of our freedom only by virtue of a common regulatory scheme.”⁴⁰ That Pettit can make sense of the ways laws interfere yet nonetheless enhance rather than impede freedom might then be taken as a unique strength of Pettit’s view of freedom as non-domination, and against my view of laws as actual interferences.

But the worry is misguided. For a law interfering in one respect is perfectly compatible with its protecting one from interference in other respects. Take the traffic-law example. Undoubtedly a law against traveling faster than 55 mph interferes with one’s freedom to drive faster than 55 mph. But such a law also protects one’s freedom to travel the roads without being hit by a speeding car, the violence of which would surely amount to a far greater interference. Therefore it is intuitive to say that even though a law against traveling faster than 55 mph takes away some amount of freedom, it nonetheless yields a considerably greater amount of freedom in return. Or, to invoke a comparison to a standard thought in ethical theory, just as an action can yield some pain but a greater amount of pleasure, thereby being utility enhancing overall by the lights of the utilitarian calculus, so too can a law interfere with one to a small degree but also yield a greater amount of noninterference—i.e., protection from interference—thereby being freedom enhancing overall.⁴¹

39 Pettit, *Just Freedom*, 66.

40 Pettit, *Just Freedom*, 89, 85.

41 Of course, precisely measuring degrees or quantities of freedom is a famously vexed issue. Berlin thought that no precise measure is possible, though he did think that one can make justifiable comparisons, such as a citizen of a democracy being more free than a subject in a monarchy (*Two Concepts of Liberty*, 43). That said, many have attempted to measure

The example is hardly unique. Consider another: a law mandating that food-service employees wash their hands after using the restroom. Surely this law interferes with the freedom of an employee to choose to not wash his or her hands after using the restroom, but the resulting loss of freedom to poisoned customers is orders of magnitude greater; if a person causes my sickness I lose the freedom to do all the things I would do were I healthy, whereas the degree of loss of freedom to the newly sanitized employee is quite minimal. So a regulation that requires sanitary standards, despite being an interference, can nonetheless, on balance, yield more freedom as noninterference overall. Note too that such laws do not require those who lose and those who gain freedom in these respects to be different people; one and the same food-service employee who loses freedom *qua* an employee due to the interference of the mandatory hand-washing policy gains a greater compensating freedom *qua* a customer at every other restaurant.

There are options as to how one might formulate the distinction here. One way might be between *prima facie* and *ultima facie* or all-things-considered interferences. Another might be between there being some overall, net, or on-balance interference—this seems to be what Kramer has in mind with his distinction between particular and overall freedom.⁴² Adjudicating between these is unnecessary, however, as the basic point is the same. A law can be an actual interference while simultaneously yielding more protection from interference and so more freedom overall. So this account can indeed explain what Pettit takes to be truths: that “regulation is not oppression,” and that “we establish most of our freedom only by virtue of a common regulatory scheme” (as quoted above). One does not require a distinct category of (non)domination to make this judgment: one need only distinguish between some and net interference, as doing

freedom more precisely; see Carter, *A Measure of Freedom*, ch. 7, for a response to Berlin's skepticism in particular. Still, and as best I can tell, there is no approach to the measurement of freedom that is wholly satisfying, agreed upon, or theory neutral. For example, measurements of freedom in terms of preferences presuppose, *contra* what might be the lesson of the contented-slave case, that freedom should be understood in terms of actual preferences; cf. Sugden, “The Metric of Opportunity.” Similarly, measurements in terms of “opportunity sets” (Pattanaik and Xu, “On Ranking Opportunity Sets in Terms of Freedom of Choice”) seem to presuppose (*contra* Taylor, “What's Wrong with Negative Liberty?”) that it is the number of opportunities, rather than their significance, that is directly relevant to the degree of freedom enjoyed. Adjudicating these complex issues is not necessary here, however; for even without a theory-neutral or agreed-upon metric, my judgments regarding traffic laws and sanitary conditions (see below) seem uncontroversial and unlikely to be overturned by appeal to a formal metric of freedom.

42 Kramer, “Liberty and Domination,” 32.

so yields the result that negative freedom can indeed be promoted rather than inhibited by laws and regulations.⁴³

4. CODA: THEORIES, EXPLANATION, AND FORMS OF GOVERNMENT

Formulating the most recent argument as I do allows me to tie together a couple of threads, as well as respond to a final objection. Though I noted that it was not a focus of this paper, I mentioned at the outset that connected to the issues I do discuss is the relationship between freedom and forms of government. In particular I noted that on Pettit's non-domination view, there is a necessary or conceptual relation between democracy and freedom, and authoritarian (dominating) governments and unfreedom. Whereas for Carter and Kramer it is only contingent and empirical that democracy promotes freedom more than, say, fascism. Recall also that I have appealed to methodological issues, often regarding explanation, and noted that debates over theories often turn on the explanation of certain data or cases, such as the datum that slavery is paradigmatic of unfreedom.

Assuming this methodology, suppose it is a datum that certain laws or regulations can enhance rather than impede freedom, such as the traffic-law example discussed above. How is this to be explained? Pettit's explanation appeals to non-domination: because citizens are not dominated when a democratic regime imposes traffic laws, one remains free even when governed by or interfered with by such laws. I offered a distinct explanation for the same datum, however: traffic laws, on balance or *ultima facie*, protect against more interference than they impose. So on my explanation there is no distinctive appeal to forms of government, nor to the question of whether the laws reflect an "alien will," or track the "avowed interests" of the citizens in question. Instead, my explanation simply appeals to which option—traffic anarchy or a system of regulations—would yield more interference, with my claim being that the former yields more interference than the latter because the regulations protect one against the interference of unruly drivers.

But perhaps one thinks this is not the whole story, and that traffic laws not taking away freedom is not the only datum to account for. For one might also think it is a datum that such laws do not take away freedom *because* they are not dominating (or reflective of an alien interest or will). Then one might think my

43 Similar remarks apply to List and Valentini's adoption of freedom as independence, which they motivate, at least in part, via its ability to explain how regulations may promote freedom—in putative contrast to the inability of freedom as noninterference ("Freedom as Independence," 1072).

explanation comes up short. More generally, one might think that people subjected to democratic laws are not dominated and therefore, or because of that, not unfree. And correlatively one might think that those subject to authoritarian or fascist laws are unfree because they are dominated. And if so, one might object that the noninterference theory I have defended cannot explain these facts.⁴⁴

But this objection is not successful, for reasons both substantive and methodological. Starting with the latter, while I accept that a theory must explain certain facts or data, some tenets or claims need not be explained. Compare: while any theory of freedom must account for the unfreedom induced by slavery, it is not the case that any theory of freedom must account for the unfreedom induced by poverty. The reason is that it is contentious whether poverty takes away freedom; while certainly one may argue (and many have argued) that poverty does take away freedom, this is better understood as an implication of a theory, or a judgment one makes as a result of a theory, rather than a theory-independent datum that a theory must account for.⁴⁵ More generally, proper methodology here requires distinguishing theory-neutral facts, data, or paradigms, such as slavery taking away freedom, from contentious (even if true) judgments, such as poverty taking away freedom. For theory-neutral facts or data are in a sense antecedent or prior to a theory, whereas a judgment generated by or justified in light of a theory is posterior.

Much the same applies to putative explanations of a datum. While certain data or theory-neutral facts do need to be explained, the claim that some datum or fact is the case *because of* a certain explanatory factor is not itself part of what needs to be explained. So assume for argument's sake that certain laws, such as traffic laws, enhance rather than impede freedom. Still, one must distinguish needing to explain this fact from the (putative) need to explain this fact obtaining *because of* non-domination. Simply put, a theory-laden or theory-based explanation of a datum is not itself among the data that any theory must account for, on pain of being explanatorily inadequate. So it is not the case that the noninterference theory I am defending must explain why certain laws are freedom enhancing *because they are not dominating*, even though I accept the burden of needing to explain why certain laws are freedom enhancing. And that is what I did above: by explaining that regulatory laws can yield more protection from interference, on balance or *ultima facie*, than would be yielded in their absence.

Still, perhaps one thinks there is an important fact (or datum) to be explained here regarding different forms of government. In particular, one might

44 My thanks to an anonymous referee for the *Journal of Ethics and Social Philosophy* for offering this objection, and encouraging this discussion.

45 Cf. note 36 above.

think there is something inherently or necessarily freedom enhancing about democracy, and freedom limiting about nondemocratic or authoritarian forms of government. And perhaps one thinks Pettit's view retains a superior ability to explain these data.

But this is not the case, for reasons closely related to those just detailed. To see this, first distinguish between laws that might obtain in a democracy *and* an authoritarian state, as opposed to laws that would only obtain in an authoritarian state. For example, I assume that democracies and authoritarian states could have similar traffic laws, while only an authoritarian state would have a law requiring worship of the supreme leader. From here, note that it cannot be taken as a theory-neutral datum that people in a democracy are more free vis-à-vis or with respect to their *traffic laws* than people in an authoritarian state are with respect to *their* traffic laws; though one could argue that there is such a difference, it is hardly self-evident or a theory-neutral datum that any theory must account for. For if we assume that the content of the laws is identical, differing only in how they were enacted (e.g., legislature versus dictate), then the effects of those laws would be indiscernible, just as two identical actions born of different motivations would be behaviorally indiscernible. So one cannot assume that there is a difference in freedom in these situations that needs to be explained. So *a fortiori* one cannot assume that there is a difference in these situations *because* of the absence or presence of domination. Such judgments are simply too theoretically loaded to be among the neutral data that any theory must account for.

On the other hand, if the *explanandum* is the datum that people are unfree when required to worship a supreme leader, say, then the noninterference conception I have defended clearly has the resources to explain this. For such a requirement is a gross interference with the beliefs, thoughts, and attitudes of every person subject to this law, *mutatis mutandis* for any other law that would exist in an authoritarian state but not a democracy. Yet even if one thinks the *best* explanation of the unfreedom here appeals to domination, recall that earlier I argued that one should construe domination as a particularly gross or insidious form of interference; some interference is so severe, persistent, and aimed at subjugation that it is fairly described as "domination." For, I argued, it is only by assuming that domination obtains when interference does not (as in the lucky-slave case) that domination appears categorically different than interference. Yet accepting the wide-scope conception of interference dissolves the opposition, as one can instead see domination as a kind of interference (the persistent or subjugating kind). Thus the wide-scope conception of interference I have defended can explain unfreedom wherever there is domination.

And this in turn allows a final point. Recall again that Pettit differs with Carter

and Kramer over the strength of relationship between slavery and (un)freedom, as well as forms of government and (un)freedom: whereas for Pettit the relationships are necessary or conceptual, for Carter and Kramer they are contingent and empirical. Though like Carter and Kramer I have defended a negative noninterference conception against Pettit, I also argued earlier, with Pettit and against Carter and Kramer, that slavery entails and so necessarily yields unfreedom—the reason being, *contra* Pettit, that a slave is necessarily interfered with by slavery. The same applies to forms of government: whereas democratic laws that protect citizens from interference *entail* noninterference and so freedom, subjugating or dominating authoritarian laws entail gross interference and so unfreedom. The relationship is not merely contingent or empirical.

5. CONCLUSION

Pettit argues that the liberal conception of freedom as noninterference is unable to explain crucial instances of freedom and unfreedom—in particular the unfreedom of slavery, and the freedom that results from certain regulatory laws. I have shown that these arguments are unsound. In particular I have argued that the scope of interference is wider than both Pettit and his critics have recognized, and that this, among other reasons, shows the lucky-slave case to be an impossibility rather than a possibility with low probability. More generally I have shown that freedom as noninterference, conjoined with a proper auxiliary assumption regarding the wide scope of actual interference, accounts for paradigm cases of unfreedom. I have also shown that freedom as noninterference has the resources to make sense of the ways in which laws and regulations may enhance rather than inhibit freedom, as well as how different forms of government may enhance or inhibit freedom. Conjoined, I have shown that the negative conception of freedom as noninterference can meet the explanatory burden that Pettit has claimed it cannot.⁴⁶

College of William and Mary
jgoldwater@wm.edu

⁴⁶ My sincere thanks to Victoria Costa, Laura Ekstrom, Christopher Freiman, Chad Vance, and two anonymous referees for the *Journal of Ethics and Social Philosophy* for helpful comments on earlier versions of this paper.

REFERENCES

- Beckman, Ludvig, and Jonas Hultin Rosenberg. "Freedom as Non-Domination and Democratic Inclusion." *Res Publica* 24, no. 2 (May 2018): 181–98.
- Benn, S. I., and W. L. Weinstein. "Being Free to Act, and Being a Free Man." *Mind* 80, no. 318 (April 1971): 194–211.
- Berlin, Isaiah. *Two Concepts of Liberty*. Oxford: Oxford University Press, 1969. Reprinted in Miller, *The Liberty Reader*, 33–57. Page references are to the reprinted edition.
- Carter, Ian. "How Are Power and Unfreedom Related?" In Laborde and Maynor, *Republicanism and Political Theory*, 58–82.
- . *A Measure of Freedom*. Oxford: Oxford University Press, 1999.
- . "Positive and Negative Liberty." *The Stanford Encyclopedia of Philosophy* (Fall 2016). <https://plato.stanford.edu/archives/fall2016/entries/liberty-positive-negative>.
- Cohen, G. A. "Capitalism, Freedom, and the Proletariat." In Miller, *The Liberty Reader*, 163–82.
- Costa, M. Victoria. "Neo-Republicanism, Freedom as Non-Domination, and Citizen Virtue." *Politics, Philosophy, and Economics* 8, no. 4 (November 2009): 401–19.
- de Bruin, Boudewijn. "Liberal and Republican Freedom." *The Journal of Political Philosophy* 17, no. 4 (December 2009): 418–39.
- Flikschuh, Katrin. *Freedom: Contemporary Liberal Perspectives*. Cambridge: Polity Press, 2007.
- Goodin, Robert E. "Folie républicaine." *Annual Review of Political Science* 6 (2003): 55–76.
- Goodin, Robert E., and Frank Jackson. "Freedom from Fear." *Philosophy and Public Affairs* 35, no. 3 (Summer 2007): 249–65.
- Harbour, Michael David. "Non-Domination and Pure Negative Liberty." *Politics, Philosophy, and Economics* 11, no. 2 (May 2012): 186–205.
- Kramer, Matthew H. "Liberty and Domination." In Laborde and Maynor, *Republicanism and Political Theory*, 31–57.
- . *The Quality of Freedom*. New York: Oxford University Press, 2003.
- Laborde, Cécile, and John Maynor, eds. *Republicanism and Political Theory*. Malden, MA: Blackwell Publishing, 2008.
- Lang, Gerald. "Invigilating Republican Liberty." *The Philosophical Quarterly* 62, no. 247 (April 2012): 273–93.
- List, Christian, and Laura Valentini. "Freedom as Independence." *Ethics* 126, no. 4 (July 2016): 1043–74.

- Miller, David, ed. *The Liberty Reader*. Boulder, CO: Paradigm Publishers, 2006.
- Pattanaik, Prasanta, and Yongsheng Xu. "On Ranking Opportunity Sets in Terms of Freedom of Choice." In *Freedom: A Philosophical Anthology*, edited by Ian Carter, Matthew H. Kramer, and Hillel Steiner, 463–68. Oxford: Wiley-Blackwell, 2007.
- Pettit, Philip. "The Instability of Freedom as Noninterference: The Case of Isaiah Berlin." *Ethics* 121, no. 4 (July 2011): 693–716.
- . *Just Freedom: A Moral Compass for a Complex World*. New York: W. W. Norton and Company, 2014.
- . "Republican Freedom: Three Axioms, Four Theorems." In Laborde and Maynor, *Republicanism and Political Theory*, 102–30.
- . *Republicanism: A Theory of Freedom and Government*. Oxford: Oxford University Press, 1997.
- Skinner, Quentin. "A Third Concept of Liberty." *Proceedings of the British Academy* 117 (2001): 237–68.
- Sugden, Robert. "The Metric of Opportunity." *Economics and Philosophy* 14, no. 2 (October 1998): 307–37.
- Taylor, Charles. "What's Wrong with Negative Liberty?" In Miller, *The Liberty Reader*, 141–62.
- Waldron, Jeremy. "Homelessness and the Issue of Freedom." *UCLA Law Review* 39, no. 1 (1991–92): 295–324.
- Wendt, Fabian. "Slaves, Prisoners, and Republican Freedom." *Res Publica* 17, no. 2 (May 2011): 175–92.

DO WE HAVE REASONS TO OBEY THE LAW?

Edmund Tweedy Flanigan

IN THIS PAPER, I shall ask whether we have content-independent moral reasons to obey the law, and I shall make some claims about what we mean when we ask this question. I shall also inquire after the strength of such reasons. In other words, I shall ask whether we ought morally to do as the law demands, because the law demands it.

This is a version of a very old question, but it is in one important way different from that question. Traditionally, we ask whether subjects are *morally obligated* to obey the law, or whether, equivalently, they have a *moral duty* to do so.¹ I believe we should instead begin with the more modest question of whether we have *moral reasons* to obey the law. This question retains the structure of the traditional question, but it is at the same time simpler and clearer, and so easier to answer. Moreover, many take reasons to contribute to obligations, so in answering the question about reasons, we may also make progress on the question about obligation.²

The answer to this question I shall defend is yes: we *do* very often have moral reasons to obey the law, *because it is the law*, in the content-independent sense. Moreover, I shall suggest that these reasons very often amount to an obligation to so act.

This answer goes against a strong current in political and legal philosophy which has led many to endorse philosophical anarchism, the family of views often expressed by some combination of the claims 1–3:³

- 1 Following common practice in this literature, I will not distinguish between *obligation* and *duty*. My discussion of the former should be read as applying as well to the latter. The best-known view that does distinguish between these concepts in this context is Rawls's. For him, obligations are interpersonally incurred (like promises, for example), whereas duties may be "natural." See Rawls, *A Theory of Justice*, 98–101, sec. 19.
- 2 I substantiate this point in slightly greater detail in the next section. Beyond this, the benefit of asking the reasons question before the obligation question will also (I hope) be demonstrated by the fruit of what follows, taken as a whole.
- 3 Matthew Noah Smith goes so far as to write that "there may be a consensus amongst moral and political philosophers that there is not today any existing obligation to obey the law"

1. Any reasons we may seem to have to do as the law demands are really just reasons to do as we ought *full stop*, independent of the law, in virtue of our ordinary moral obligations; or
2. Whatever conditions would obligate us to do as the law demands are not met, and maybe could not be met, by the law; or
3. Any such apparent reasons are merely *prudential* reasons to act so as to avoid being fined or punished by the state.

The law, on any of these views, is not morally significant.⁴

Separately, some have been recently convinced that there can be no content-independent reasons to obey the law, or at least that no successful account of what such reasons might amount to has yet been given.⁵ This separate conclusion only strengthens the appeal of the philosophical anarchist's claim, by undermining the very possibility of having a reason to do as the law demands because it is what the law demands.

I believe we can answer both of these skeptical challenges. Content independence is not as mysterious as it has often been made to seem. On the view I propose, when we talk of content independence, we are making claims about *grounding*. When we claim that the law provides content-independent reasons for its subjects to ϕ , we are claiming that there is a distinctive property of the law which grounds a moral reason to ϕ , which is another way of claiming simply that the law's distinctive properties are *morally significant*.

When we understand content independence in this way, it becomes easier to see that the law very plausibly does provide content-independent reasons to do as it demands. We can also see that such reasons may often combine, sometimes with nonlegal reasons and sometimes on their own, to amount to an *obligation* to obey the law. While my remarks on this point must remain schematic—whether we are in fact obligated by the law depends on further commitments regarding the status and normative force of various candidate properties of the law, regarding the normative circumstances of particular subjects, as well as regarding com-

(“Political Obligation and the Self”). Similar claims may be found in Edmundson, “State of the Art”; Gur, “Are Legal Rules Content-Independent Reasons?”; Klosko, “Are Political Obligations Content Independent?”

- 4 I do not here address Wolff's early anarchist view, which focuses on the agential costs to following an authority's directive, nor Smith's recent defense of a related view. To address these views, distinctive as they are, would require a separate paper. As Smith notes, moreover, that view “has never had much traction” in the literature (though his paper is an attempt to revive it). See Wolff, *In Defense of Anarchism*; Smith, “Political Obligation and the Self.”
- 5 See Markwick, “Law and Content-Independent Reasons” and “Independent of Content.” Sciaraffa argues for a similar conclusion (“On Content-Independent Reasons”).

peting conceptions of the concept of obligation—we can nevertheless conclude that, very plausibly, the law often succeeds in morally obligating us. The anarchist's position is thus importantly undermined.

1. PRELIMINARIES

It may help to begin by first offering some definitions and clarifications, if only because the literature on political obligation has suffered, in my view, from some unclarity about reasons. I follow Scanlon and Parfit in using the “purely” or “genuinely” normative concept of a reason, according to which a reason to ϕ may be helpfully redescribed as a fact that counts in favor of ϕ -ing.⁶ To have at least one reason to ϕ is the same as having “some reason,” or simply “reasons” to ϕ , though we may have some reason to ϕ even when there is some other act that we ought to do instead, because the reasons favoring that act are stronger than our reasons to ϕ .⁷ When the balance of reasons counts decisively in favor of our ϕ -ing, in the sense that our reasons to ϕ outweigh any competing reasons not to ϕ , or to do some other act instead, we can say that we have “decisive reason” to ϕ , which is one way of saying simply that we ought to ϕ .⁸ If our reasons are such that we may permissibly ϕ but are not required to ϕ , we say that we have “sufficient reason” to ϕ .

When I speak of someone's “having” a reason to ϕ , I do not mean to imply anything about this person's own awareness of her reasons, nor about her motivational states.⁹ In the way I use the term, a person's having a reason to ϕ is just the same as there being a reason for her to ϕ , which is just the same as there being some fact that counts in favor of her ϕ -ing. Similarly, when I say that some fact “gives” or “provides” us with a reason to ϕ , I mean only that that fact is a reason for us to ϕ , by counting in favor of our ϕ -ing. By extension, when I say that the law gives or provides us with a reason to ϕ , I mean that the fact that the law demands that we ϕ is a reason for us to ϕ , or counts in favor of our ϕ -ing.¹⁰ In

6 Scanlon, *Being Realistic about Reasons*, 1–15; Parfit, *On What Matters*, 1:31–37.

7 Some refer to these as *pro tanto* reasons, as a way of indicating that these reasons can weigh together, outweigh, and be outweighed by other reasons. For my purposes, writing “*pro tanto*” in front of “reason” does not add anything, as all of the reasons I discuss can weigh together, outweigh, and can be outweighed by other reasons.

8 There are other senses of “ought,” but I shall stick to the decisive reason-implying sense here.

9 The exception to this is when I consider the view, defended by Hart and Raz, that our reasons to obey the law require that we act *for* certain reasons and not others. See section 3.2 below.

10 It may be natural to talk of some people, or even the law, “giving reasons” to others, but such talk is often misleading, and I shall avoid it. Facts supply reasons, and the only helpful

this paper I shall discuss only ideal cases, so there will be no need to distinguish between reasons for action and the merely apparent or contextually normative reasons for action that we may be said to have, or be aware of, in virtue of our beliefs about which facts count in favor of which acts. Only in these nonideal cases is it useful, I believe, to talk of someone's "having a reason" in this other sense, and similarly, only in nonideal cases is it useful to talk of something "giving" a reason to someone in the corresponding sense.

Additionally, it is important to say that I take myself to be discussing *moral* reasons, as distinct from merely prudential reasons, epistemic reasons, reasons of rationality, and so on. I do not take a view on the issue of what makes our moral reasons distinct from other kinds of normative reasons; I claim only that there is a useful distinction to be drawn.

Another point of unclarity in the literature has been regarding the nature of *obligation*. As mentioned already, we are traditionally confronted with what I shall call the "obligation question," which asks,

Are subjects *obligated* to do as the law demands, in virtue of it so demanding?

However, rather than begin with the obligation question, I suggest that we first ask the more modest "reasons question":

Do subjects have *reasons* to do as the law demands, because the law demands it?

This is for several reasons. First, because the concept of obligation, particularly in this context, is insufficiently clear. Its various conceptions are not normally well distinguished, yet what we mean by "S is obligated to ϕ " of course bears importantly on what we should think about whether subjects are obligated to obey the law.¹¹ Second, the concept of a reason is simpler, as well as (arguably) more

sense in which people may give reasons to others is by helping create (as by promising or commanding) or by calling attention to (as by pointing out) such facts. On this point, see Enoch, "Reason-Giving and the Law."

11 For instance, a common thought is that for S to be under a moral obligation to ϕ is for it to be wrong for S not to ϕ . While this may be an attractive understanding of moral obligation in general, it is less clearly helpful in the context at hand, not least because it is unclear who or what would be wronged by a failure to obey the law. By contrast, Green analyzes obligations in this context as requirements with which subjects are "*bound* to conform," where the notion of being bound is explained as being "nonoptional" or compulsory. See Green, "Legal Obligation and Authority" (emphasis in original). He is here following Hart, who makes similar remarks in "Legal and Moral Obligation." This is intuitively closer to what I believe most theorists in this literature have in mind, although it is clearly stronger than the

fundamental, than the concept of an obligation. We can thus more clearly know what we are asking when we ask the reasons question, and so we can more clearly know whether we have an answer. Third, and relatedly, because it is widely thought (and I believe) that reasons *contribute* to obligations. Precisely how they do so is a matter of debate, though we need not make any more specific commitments here than that they *do*. If it is true that reasons contribute to obligations, then in answering the reasons question, we have also (at least partly) answered the obligation question.

Because I ultimately wish to discuss philosophical anarchism, whose standard formulation denies that subjects are morally obligated to obey the law, I will not be able to entirely avoid obligation talk. For our purposes, I shall use “normally decisive reason” as a moderately ecumenical analysis of the concept of obligation. This is not because I take it to be a particularly good analysis of what obligation *is* (it is not), but rather because it seems extensionally compatible with many reasonable ways of speaking about obligation and duty, such as when Ross discusses our “*prima facie* duties” to act in certain ways, and also with the idea that obligations are those acts which are, in view of the balance of reasons, morally required.¹² Still, my use is a substantive commitment, and so alternative conceptions of obligation may lead to disagreement with my claims later in the paper about our reasons to reject philosophical anarchism. Since I know of no anarchist position that understands obligation in a way that is incompatible with my commitment here, it would fall to the anarchist to develop such a position.

Finally, it should be borne in mind throughout that we must always be careful to distinguish some reason for action [*r*] from a “summary reason” given by a set of reasons for that action that includes [*r*] as one member among others of the set, lest we double count the reasons.¹³ Thus, to borrow Parfit’s example, the fact that some medicine is the cheapest and most effective may make it the best medicine, but when we talk of the reasons for some person to take this medicine, we would make a mistake if we claim that this person has *three* reasons to do so: that it is the cheapest, the most effective, *and* the best.

previous conception. Moreover, these further notions (of being “bound,” “nonoptionality,” and so on) are hardly more perspicuous than the original.

12 Edmundson makes a similar claim, writing that the duty to obey the law is regarded “as one that is ordinarily decisive” despite being “subject to being defeated or outweighed by countervailing moral considerations.” See “State of the Art,” 215–16.

13 Throughout the paper, I will use brackets to indicate facts: “[*r*]” means “the fact that *r*.”

2. CONTENT INDEPENDENCE

We can now consider the idea of the *content independence* of certain reasons, including reasons given by the law. Roughly, the idea is that a reason is content independent if, as Hart puts it, it is “intended to function as a reason independently of the nature or character of the actions to be done”; or, as Raz puts it, if “there is no direct connection between the reason and the action for which it is a reason.”¹⁴ In the case of content-independent “legal reasons”—the reasons given by acts being prohibited, permitted, or required by the law—we say that such reasons are to do as the law demands, *whatever* the law demands, no matter the moral, rational, or perhaps even legal merits of what is demanded.¹⁵ Taken another way, the idea of content independence is the thing that we mean by “because it is the law” when we discuss the claim that we have reasons to obey the law *because it is the law*.

It will help to begin by discussing a recent skeptical challenge. The legal philosopher Paul Markwick has rightly questioned the idea that all reasons are, as such, either content dependent or content independent. Reasons are, I have claimed, facts that count in favor of actions. On this understanding, it is mysterious what it would mean to claim that some reasons bear the fundamental property of content dependence while others bear the fundamental property of content independence. What is the content of a reason, other than the fact that constitutes it, or that fact’s propositional content? How could any reason be independent of that? And what would we add to the claim that [*r*] is a reason for someone to ϕ by making the further claim that that reason is dependent upon or independent of *r*? It is not clear that such a claim would even make sense.

Most of those who claim content independence for legal reasons do not take the content in question to be the content of the reason *per se* but rather the thing that the reason is a reason to do. Thus, for some reason to ϕ , the content of that reason is ϕ , or ϕ -ing.¹⁶ It is this that Markwick has in mind when he argues that legal reasons are not distinctively content independent. As I shall now argue, I think Markwick is correct in this view but also partly misled by his argument, so

14 Hart, *Essays on Bentham*, 254; Raz, *The Morality of Freedom*, 35–37. It has often been remarked, *pace* Hart, that intention seems largely beside the point.

15 The notion is typically understood to allow limiting cases, such as when the law’s demands are grossly immoral or unjust, or perhaps when the law’s demands are too demanding. I shall ignore such limiting cases—while acknowledging their deep importance in other contexts—for the purposes of this discussion.

16 Strictly speaking, it thus seems both more accurate and clearer to refer to ϕ as the *object* of *S*’s reason to ϕ rather than its *content*, but it is probably too late to correct that particular error.

that he rejects the notion of the content independence of legal reasons altogether as uninteresting or uninformative. I do think legal reasons are not *distinctively* content independent, but I do not think the notion is therefore uninteresting or uninformative. Rather, I think that by better understanding the sense in which legal reasons are often claimed to be content independent, and by seeing how such reasons are not so *unlike* other kinds of normative reasons, we can better see how we might have reasons to obey the law *because it is the law*.

In one paper, Markwick considers the following candidate condition for content independence of reasons:

If ϕ -ing's F -ness is a reason to ϕ , this reason is content independent if and only if for any other act-type μ , there would be a reason to μ if F were a property of μ -ing.¹⁷

We can restate this condition briefly as the claim that a reason to act is content independent just in case the reason is given by some property of the act such that, if another act had that property, it would also provide a reason to so act.

This seems, at first glance, like a good account of content independence. If some act has the property of being required by the law, for instance, and if having this property provides a reason to so act, then there will be a like reason to do any other act that also has the property of being required by the law. The reason given by the fact that an act is required by the law, then, would appear to be a content-independent reason.

Markwick points out, however, that this condition seems to capture too many reason types. Many acts are morally required, for example, and thus share the property of being required by morality. The reasons given by the fact that these acts are morally required thereby meet the condition above for content independence. And yet moral reasons are not typically taken to be content independent. They are, rather, commonly taken to be content *dependent*. We can also ask, if moral reasons are content independent, then which reasons are content dependent? The same point could be made about several other properties typically not thought to confer content independence upon the reasons they provide. Markwick gives two examples: the property of causing unnecessary suffering and the property of maximizing utility. Take the property of causing unnecessary suffering. That an act bears this property is a reason not to do it and would be a reason

¹⁷ Markwick, "Law and Content-Independent Reasons," 582. In this paper and in "Independent of Content," Markwick often uses the phrase "a reason" to ϕ to mean sufficient or decisive reason to ϕ . As I shall argue, however, it is much easier to argue that the law does not in all cases provide, on its own, sufficient reason to do as it demands, than to argue that it does not provide *a reason*, or *any reasons*, to so act. We should consider the latter claim about reasons first, and only then move on to stronger claims about sufficient and decisive reasons.

not to do any other act bearing this property. Such a reason would thus meet the condition above for content independence. Yet such a reason, as Markwick notes, is commonly taken to be a clear example of a content-*dependent* reason. Or take the property of maximizing utility. According to act utilitarianism, an act is required if and only if it maximizes utility, no matter any other features of the act (e.g., that doing it would break some promise, violate some people's rights, etc.). In other words, we might say, all and only utility-maximizing acts are required, *regardless of their content*. Yet surely, Markwick claims, no act utilitarian would claim that all such acts are required by content-independent reasons. For again, the question could be asked, if these reasons are content independent, then which reasons are content *dependent*? The objection, in brief, is that the fact that it is unclear which reasons might be content dependent casts doubt upon the viability and usefulness of the distinction between content dependence and independence, and further that it is "unclear how content independence is a property which distinguishes legal reasons in particular from reasons in general."¹⁸ If no reasons, or few reasons, are content dependent, or if we cannot use the property to distinguish legal reasons from reasons in general, we might urge along with Markwick that we give up on talk of content independence as an important feature of legal reasons altogether, as uninteresting or uninformative.

Part of the answer to Markwick's challenge is to concede that content independence is not a *distinctive* property of legal reasons but to maintain that, when we claim of some reasons, including legal reasons, that they are content independent, we are making a claim that is nevertheless both interesting and informative. This is because to claim that

some property of an act gives us a reason to do that act, and gives this reason the property of being content independent

is, in my view, to claim nothing more than that

this property of the act is normative, in the sense that an act's having it gives us a reason to do that act regardless of any other facts about the act.

It is interesting and informative to claim of some reasons that they are content independent simply because it is interesting and informative to claim of some properties of acts that they are normative, in the sense that they give us reasons to do those acts. It is both interesting, and would be highly informative if true, to claim that an act's maximizing utility gives us a reason to do that act, as the act utilitarian's main thesis claims. If we could truly make a similar claim of many other properties, this too would be highly informative and interesting. Such

18 Markwick, "Law and Content-Independent Reasons," 592.

properties include the property of being loved by the gods, the property of being required by the king, the property of being an act whose maxim everyone could will to be a universal law of nature, and many others, including the property of being demanded by the law. It would be informative and interesting if the property of being demanded by the law were normative, in the sense that an act's having this property provided a reason to do that act *whatever this act may be*. This is why, even conceding that content independence is not a distinctive feature of legal reasons, we may nevertheless claim that it would be interesting and informative if some legal reasons bore that property.¹⁹

It may seem, I should acknowledge now, that I have already conceded too much. If content independence is not a distinctive property of legal reasons, or if it is just another way of making the obvious claim that some reasons are given by normative properties of acts, then it may be hard to believe that I am indeed making a claim that is interesting or informative. But an important part of my thesis is that legal reasons are *not* as unlike other normative reasons as is commonly believed, and that when we claim that legal reasons are content independent, we are (at least) tacitly committing ourselves to this conclusion. Furthermore, I believe that understanding this can help to make sense of the ways in which the law may in fact be a source of genuinely normative reasons for action, such that we may truly claim that we have reasons to obey the law *because it is the law*.

19 An anonymous reviewer points out that Markwick's challenge might be thought to be directed at a straw man, in light of Rawls's famous distinction between the justification of a *practice*, following the rule-utilitarian maxim, and the justification of an *act falling under a practice*, following the act-utilitarian maxim. See Rawls, "Two Concepts of Rules." On Rawls's view, legal reasons pertain principally to the justification of practices and should therefore be construed as rule utilitarian in nature, rendering Markwick's act-utilitarian examples inapt. I take the point to be friendly: if the answer to the question (say) "ought I to obey my promise to ϕ ?" should be given in terms of my reasons to obey my promises rather than my reasons to ϕ , then there is a clear element of content independence to those reasons. If my legal reasons are generally of this nature, then the same will be true of them. That this characteristic may be shared by legal reasons and other rule-utilitarian reasons—and especially if we think the character of *morality in general* is rule utilitarian (or rule consequentialist) in nature—would not, on my view, show the property of content independence to be uninteresting or uninformative with respect to our reasons to obey the law.

The same reviewer notes that on Rawls's view, asking about the justification of a *practice* introduces an important opacity regarding the justification of acts falling under that practice, at least when those acts are constituted by the practice. The thought is that (for instance) if I have promised to ϕ , the fact that I ought to obey my promises renders inappropriate further inquiry regarding the question of whether I ought to ϕ . I take the view I propose to be compatible with various thoughts about "opaque" or "excluded" reasons, a point I discuss extensively in section 3.2 below.

To see how this is so, it will help now to more fully explain my view of content independence. On my view, we should understand claims of content independence as grounding claims.²⁰ When we claim that some fact [*p*] grounds another fact [*q*], we can equivalently claim that

[*p*] makes [*q*] true, or makes it the case that *q*;

[*q*] is true, or holds, in virtue of [*p*];

[*q*] depends on [*p*];

[*q*] holds because of [*p*].

The grounding relation is not, it is worth emphasizing, a causal relation, nor the supervenience relation, nor is it the same as specifying the necessary conditions for some fact to hold—though instances of these other relations may sometimes coincide with instances of the grounding relation. It is, rather, the relation of one fact's making the case another fact and thereby noncausally explaining that fact. It is, appropriately to the current discussion, a dependency relation.

The relation, though difficult to define, is very familiar in normative theorizing. Consider these examples:

1. Locke writes that “this original Law of Nature for the beginning of Property, in what was before common, still takes place; and *by virtue thereof*, what Fish any one catches in the Ocean, that great and still remaining Common of Mankind . . . is by the Labour that removes it out of that common state Nature left it in, made his Property who takes that pains about it.”²¹
2. Ross asks “What *makes* acts right?” and answers that “the ground of the actual rightness of [an] act is that, of all acts possible for the agent in the circumstances, it is that whose *prima facie* rightness in the respects in which it is *prima facie* right most outweighs its *prima facie* wrongness in any respects in which it is *prima facie* wrong.”²²
3. Rawls writes that the principles generated in the original position “must hold for everyone *in virtue of* their being moral persons” and that the basis of equality lies in “the features of human beings *in virtue of which* they are to be treated in accordance with the principles of justice.”²³

20 See *inter alia* Berker, “The Unity of Grounding”; Fine, “Guide to Ground”; Rosen, “Metaphysical Dependence.”

21 Locke, *Two Treatises of Government*, ch. 5, sec. 29 (emphasis added).

22 Ross, *The Right and the Good*, 46 (emphasis added on “makes”).

23 Rawls, *A Theory of Justice*, 114, 441 (emphasis added).

4. Cohen discusses the “Pareto claim” that “inequality is indeed just when *and because* it has the particular consequence that it causes everyone to be better off.”²⁴
5. Some people claim that a person deserves some treatment *because of* this person’s prior acts or bad character.

The grounding relation is normatively indispensable; there are very many other such examples all around us.²⁵

The claim that legal reasons are content independent is a claim about which features of the law can make it the case that we have reasons to do as it demands. When we consider the claim that we ought to obey the law *because it is the law*, the “because” is the because of grounding. When we consider the claim that it is in virtue of being *against the law* that some act is wrong, rather than due to the “merits of the act itself,” we are considering the claim that some feature of the law that is not also a feature of the act demanded by the law is what makes the act wrongful, or what *grounds* its wrongness.

Put most simply, the claim is that

[The law demands that $S \phi$] grounds [S has a reason to ϕ].

However, to claim just that [The law demands that $S \phi$] grounds [S has a reason to ϕ] may be misleading, for demands do not by themselves ground reasons for action. Rather, it is only in combination with the facts that legitimate those demands that they may do this. If the law’s demands can ground reasons for action, there must be some further feature of the law that gives its demands this force.

In many cases, it is worth noting, a legitimate demand may play no part at all in grounding a reason for action. This may be easiest to see by considering a basic case of promising. Suppose I have made you a promise that I will ϕ , and that you subsequently demand that I fulfill my promise. We could not then claim that

[You demand that I fulfill my promise to ϕ] grounds [I have a reason to ϕ]

because it is not your demand but my promise which grounds the obligation: I

²⁴ Cohen, *Rescuing Justice and Equality*, 89 (emphasis added).

²⁵ I have given only a few easy examples. A search of works in moral and political philosophy for the terms “in virtue of,” “makes it the case that,” and “when and because” will give a sense of just how often the normative grounding relation is appealed to. Nor is this a distinctively contemporary or even modern phenomenon. When Plato’s Euthyphro asks whether acts are pious because the gods love them, or loved by the gods because they are pious, he is asking a question about normative grounding. (In “No Church in the Wild,” Jay-Z asks the same question poetically: “Is Pious pious ‘cause God loves pious?” He too is asking a question about grounding.)

would have a reason to fulfill my promise regardless of whether you demanded that I do so or not. More fully, it is in virtue of my having made a promise to you both that you may make a genuinely normative demand of me that I fulfill my promise *and* that I have a reason to do so.

By contrast, some demands may seem to *by themselves* ground reasons for action. Certain demands by those we love, for example, or to whom we otherwise have special obligations, may seem to be clear cases of this kind. If your child demands love and attention, for instance, it may seem that

[Your child demands love and attention] grounds [You have a reason to give your child love and attention].

But even in this case, it would be better to claim that your child's demand *together with* your special obligations to your child ground your reason for action. To claim only that your child's demand grounds your reason for action would be misleading, and to claim that your child's demand *by itself* grounds your reason for action would be false.

Similar remarks apply to the law's demands on us. In order to make plausible the claim that the law's demands ground reasons for action, we need to identify some property or properties of the law in virtue of which its demands are genuinely normative. In other words, it is only in virtue of some normative property or properties of the law *together with* the fact that the law makes specific demands of us that we may come to have reasons to do as it demands. We should therefore consider the revised claim that

[The law has some properties $\{P\}$] and [The law demands that $S \phi$] together ground [S has a reason to ϕ].

Note that we are now discussing the possibility of several facts together grounding a single fact. To understand this, it may help to make explicit my assumption that not all grounding relations between one fact and another are relations of *full grounding*; many are relations of *partial grounding*. One fact [p] partially grounds another [q] when [p] *helps* make [q] true, or *helps* make it the case that [q], or when [q] holds *partly* in virtue of, or *partly* because of, [p]. The fact that S has a reason to ϕ , for example, would *partially ground* the fact that S ought to ϕ , by contributing to the set of reasons for S to ϕ . The set of reasons such that S has more reason to ϕ than to do any alternative act *fully grounds* the fact that S has decisive reason to ϕ , by fully making it the case that S ought to ϕ .²⁶ The claim I wish to make may indeed be more fully stated as the claim that

26 There are also many nonnormative examples of the distinction between full and partial

Each of the facts [The law has some properties $\{P\}$] and [The law demands that $S \phi$] partially ground, and together fully ground, [S has a reason to ϕ].²⁷

This claim about the content independence of legal reasons can be made in other, equivalent ways. The grounding claim made by claims of content independence of the law is also a claim about what *does not fully ground* a person's reason for action. Namely, it is the claim that the fact that a law demands that $S \phi$, and indeed all facts about ϕ -ing, *do not fully ground* the reason given by the law for S to ϕ . Or in other words, the legal reason for S to ϕ is *independent* of facts about ϕ -ing, where "independent" is to be understood as a negative grounding claim. The *positive* grounding claim is that when S has a content-independent reason to obey the law's demand that $S \phi$, this reason is at least partially grounded by some fact or facts about the law *that are not also facts about ϕ -ing*. These facts are about the normative properties of the law in general, as distinct from any particular action the law requires or forbids.

Understanding the claim that legal reasons provide content-independent reasons for action as a claim about grounding helps make sense of a number of cases in which, intuitively, it may be unclear whether the law provides a genuinely normative, content-independent moral reason to do as it demands. Imagine, for instance, a society in which the law is merely a codification of morality, such as that of the Israelite tribe under Moses: Moses's tablets, we can imagine, were a codification of the independently normative moral truths given to the Israelites by God. We might then imagine one Israelite appealing to another, more murderously inclined Israelite that the tablets forbid killing. "You must not kill, because it is against the law," the one might say. What should we make of the first Israelite's appeal to the second?

Let us grant that the second Israelite in fact has a reason not to kill. If the first Israelite's appeal is meant to make a claim about what grounds the second Israelite's reason, then that claim is false, since the fact that "THOU SHALT NOT

grounding, which may be clearer: [The apple is golden and delicious] is fully grounded by [The apple is golden] and [The apple is delicious] *together*, and partially grounded by each of those facts *separately*.

27 One of the law's properties is of course that it demands that $S \phi$, and another is, presumably, that it demands that a certain set of people including but not solely consisting of $S \phi$. The first would merely repeat [The law demands that $S \phi$], since to make this claim is to make the claim that the law has the property of demanding that $S \phi$, so would not in that case constitute additional partial grounding for [S has a reason to ϕ]. As for the second, to claim that the wider [The law demands that a set of people $\{S, \dots\} \phi$] together with its instantiation [The law demands that $S \phi$] grounds (if indeed it does ground) [S has a reason to ϕ] is not importantly different from simply claiming that the instantiation alone can do this work.

KILL” is written on Moses’s tablets certainly does not ground the second Israelite’s reason not to kill—it is rather the moral prohibition on killing, given to the Israelites by God and recorded by Moses on the tablets, that does this. Just as the wind’s happening to spell out “THOU SHALT NOT KILL” in the sand would not itself be a reason not to kill, etchings of God’s moral law in stone would not themselves provide reasons for action. Only speaking loosely may we claim that an act’s being prohibited by Moses’s tablets is a reason for the Israelites not to do it, and when we make such a claim, it must be either false or else, more charitably, be a shorthand way of making the more accurate claim that the act is prohibited by God. This is because, in the case we are imagining, God’s prohibition is genuinely normative, whereas being codified on Moses’s tablets, or being written by the wind in the sand, is not.

More generally, when demanding that for each morally required act, *S* do that act, we would make a mistake by claiming that

[The law codifies morality] and [The law demands that *S* ϕ] together ground [*S* has a reason to ϕ].

For it would be in virtue of ϕ -ing’s being morally required, rather than in virtue of the law’s demanding or codifying that *S* ϕ , that *S* has a reason to ϕ . These further facts about the law would add nothing to the normative grounds for *S*’s ϕ -ing, which is to say would not help to make it the case that *S* has a reason to ϕ . In the same way, if I told you truly that you ought morally to do some act, and even if I *always* told you truly what the thing was you ought to do, it would be in virtue of this act’s being a moral requirement, rather than in virtue of my telling you so, that you ought to do it. In this way, the law cannot be said to provide content-independent reasons for action in cases in which it is merely a codification of more fundamental normative facts.

Similarly, we can, by understanding content independence as an idea about normative grounding, better understand the sense in which Markwick rightly claims that legal reasons are not *distinctively* content independent. The claim that legal reasons are content independent is no more than the claim that some *legal* property of these reasons, such as (say) that it was passed by a democratic assembly, or that it solves some coordination problem—rather than something about what it is these legal reasons are reasons to do—is what *makes it the case* that they are genuinely normative for those to whom they apply; just as, according to one widely held view, moral reasons are made genuinely normative not by facts about what they are reasons to do but by facts about morality, such as that the act for which the reason counts in favor would maximize utility, or is an act that no one could reasonably reject, or is an act that is in conformity with a

maxim that could be willed to be a universal law, and so on. But we can also in this way understand why the further worry that, by failing to be distinctive, the content independence of legal reasons may be uninformative or uninteresting, is misplaced. For the claim that legal reasons are content independent is the important and substantive claim that the law, like morality, can be a source of genuinely normative moral reasons for action, rather than merely a way of calling attention to reasons whose real normative force lies elsewhere.

3. OBJECTIONS

It is worth pausing now to consider some objections against the view I have proposed. First, I will consider several versions of the objection that the view fails on its own terms, since all moral reasons are ultimately grounded in facts about morality rather than the law. By explaining how this objection fails, the distinctive character of the grounding view of content independence will become clearer. Second, I will address the objection that by analyzing content independence in terms of the grounds of reasons to obey the law, I have lost (or worse, am unable to accommodate) a distinctive and important feature of our reasons to obey the law, namely, the opacity of such reasons to those subject to them. In considering these objections, I will compare my view with well-known views represented by Hart, Raz, and Rawls.

3.1. *Internal Objections*

To begin, it may be objected that *any* moral force the law has must be had in virtue of some prior moral facts, and that in view of this, no obligation to do as the law demands may be said to hold in virtue of facts about the law. This objection may take two forms. On the first, the complaint is that the law is, like Moses's tablets, a mere codification of some other normative facts, and that thus we may not properly claim that it is *the law* which grounds our reasons to do as it demands.

To answer this first version of the objection, it is important to understand the sense in which one fact's grounding another provides a noncausal *explanation* for the second fact. [The apple is golden] and [The apple is delicious] together ground *and thereby explain* [The apple is golden and delicious]. Similarly, according to act utilitarians, the claim that [The act would maximize utility] grounds *and thereby explains* [The act is required]. By contrast, if we imagine some person who always speaks truly, [This person says that the apple is golden and delicious] *would not* ground [The apple is golden and delicious], because this person's claim, though true, would *not explain* [This apple is golden and

delicious]. Nor would, for the act utilitarian, [This person says that the act is required] ground or explain [The act is required]. To *explain* in this noncausal way seems to be part of what it is to *ground*.

Now compare two similar cases. In the first, suppose that some king's dictates are independently normatively binding on his subjects, and suppose further that those dictates are published in a book of codes. In one sense, we could plausibly claim that [This act is prohibited by the codebook] grounds [This act must not be done], but only insofar as [This act is prohibited by the codebook] refers not to the physical book but to the abstract collection of dictates recorded there, so that [This act is prohibited by the codebook] is a shorthand for [This act is prohibited by the king's dictates]. This is like the way in which we can only truly claim that an act's being prohibited by Moses's tablets grounds our reasons not to do this act if we more fully mean that it is God's prohibition, which the tablets record, that grounds our reason not to do this act. In this way, there is an important sense in which we may properly say that an act's being prohibited by the codebook noncausally *explains* the fact that we must not do this act.

In the second case, suppose that some unofficial observer privately records the king's dictates in a notebook. In contrast with the first case, we could not then plausibly claim that [This act is prohibited by the notebook] grounds [This act must not be done], because the notebook is a mere private record of the king's dictates, and [This act is prohibited by the notebook] could not plausibly be a shorthand for the claim that the act is prohibited by the king. It is more like the wind writing "THOU SHALT NOT KILL" in the sand, or like, when I truly tell you that you ought not to do some act, it is not my telling you so that explains the fact that you ought not to do it. There is no sense in which being prohibited by the notebook noncausally explains the fact that some act must not be done.

When we claim that

[The law has some properties $\{P\}$] and [The law demands that $S \phi$] together ground [S has a reason to ϕ],

we are, as in the first case, making a noncausal explanatory claim. There will be further facts which ground the fact that the law has such properties, as well as, more pertinently, further facts about what makes those properties give the law normative force. If there are such properties of the law, however, which together with its demands ground reasons for us to obey it, we may properly say that those facts ground reasons for action, and we may properly refer to such reasons as reasons to obey the law *because it is the law*.

A second version of this objection argues that, when we claim that some properties of the law help ground a reason for us to do as it demands, we ought

instead to claim that it is *those* properties, or whatever grounds *them*, that ground our reasons to do as the law demands, rather than the law itself. If some facts about the democratic origins of a law, for instance, are what make it the case that we have a reason to do as it demands, then, according to this objection, we should say that it is those origins, rather than the law itself, which give us such a reason. We should, on this objection, follow the normative grounding “all the way back,” and then make any grounding claims in terms of those fundamental normative grounds.

One answer to this objection is to agree that we may often be able to make grounding claims in terms of other, more fundamental grounds by following the chain of grounding “back,” but to deny that such a move is always better. Indeed, it may on many occasions be worse to do this. When claiming that

[The apple is golden] and [The apple is delicious] together ground [The apple is golden and delicious]

it may not help, or it may be unnecessary, to reformulate the claim in terms of the very many further facts which ground [The apple is golden] and [The apple is delicious] separately. The same may be said of the facts that ground our reasons to obey the law.

Another answer to this objection is to remark that, even if some property of the law, such as its democratic origins, is what *helps* make it the case that we have a reason to do as the law demands, the fact that the law has such a property could not on its own, or fully, ground such a reason. In other words, although we might make the specific claim that

[The law’s origins are democratic] *partially* grounds [S has a reason to ϕ],

it would be misleading to claim simply that [The law’s origins are democratic] grounds [S has a reason to ϕ] because it is not only the fact that the law has this property, but also the fact that the law demands that S ϕ —that is, S as a specific subject and ϕ as a specific act—which together *fully* ground the fact that S has a reason to ϕ . Alternatively, though we might, in the case we are imagining, truly claim that

[The law’s origins are democratic] fully grounds [Subjects of the law have a reason to obey the law],

we would, in order to make a specific claim about what grounds S’s reason to ϕ , need to claim that

[The law's origins are democratic] and [The law demands that $S \phi$] together ground [S has a reason to ϕ].

We could not then, as this version of the objection urges, make the grounding claim only in terms of some fact or set of facts about the law's properties. And we should therefore *not* say that it is the law's origins, rather than the law itself, which give us a reason to do as it demands.

It should be said too that when considering which properties $\{P\}$ may help ground our reasons to do as the law demands, not just any properties will do. For, as I have argued, the property that the law codifies morality could not partly ground a reason for action. The properties in question must be, rather, *distinctive properties of the law* in order to be able to help ground a reason to obey the law, in the way that, for example, we might call some law's democratic origin, or its being part of a certain kind of fair system of social cooperation, or its being issued by a law-giving body whose authority was consented to, such distinctive properties.²⁸ While we could make the grounding claim only in these terms, we could also usefully summarize this grounding claim as a claim about why we have reasons to obey the law because it is the law. In other words, we may call the reasons grounded in these ways *legal reasons*, and we may claim that learning we have such reasons tells us something important about the normativity of *the law* rather than merely about other familiar sources of normativity. If we learn that democratic lawmaking, or legal fair play, or consent to the law can help ground such reasons, we learn something not only about the normativity of democracy, fairness, or consent, but about the moral force of the law itself.

3.2. *External Objections*

We can turn next to a different kind of objection, which takes issue not with what is entailed by the view proposed here but with what it may seem to lack. In classic discussions of the idea, content independence is typically mentioned in the same breath as another property, which Hart calls "peremptoriness," Raz calls "preemptiveness," and Rawls, though he does not give it a name, seems to have had in mind in discussing what he calls "rules of practice."²⁹ The thought uniting these discussions (which I will consider in more detail shortly) is that

28 We can also see in this way how some putatively or apparently normative properties of the law might fail to be truly normative. If a democratic lawmaking process, and the laws it produces, are justified because those laws are more likely to accurately reflect underlying moral demands, this may ground a reason to *believe* that we ought to do as the law demands, but it could not itself ground a further reason to *so act*.

29 The notions are distinct, though they are similar enough to mention as a group. I will discuss some of the differences between them below. See Hart, *Essays on Bentham*, ch. 10; Raz,

content-independent reasons to obey the law often render certain other considerations *opaque* to those subject to them—normally, the underlying considerations that justify such reasons, or other reasons for or against the thing the reason is a reason to do—by excluding those considerations from a subject's practical deliberation. By contrast, I have suggested that we may *understand* the question about the law's force simply in terms of its distinctive normative properties, and that we may *answer* that question, or begin to, by thinking about the reasons grounded by those properties. This suggests a picture of the relationship between subjects and the law that is *transparent* with respect to the grounds and strength of the law's normative force. It may thus be objected that the view proposed here is incompatible with the law's *opacity*.³⁰

As before, we can helpfully distinguish several more precise versions of this objection. The first takes as its inspiration Hart's and Raz's famous discussions of authoritative reasons, which they argue are both content independent *and* opaque.

Hart writes (endorsing a view he attributes to Hobbes) that

[a] commander characteristically intends his hearer to take the commander's will instead of his own as a guide to action and so to take it in place of any deliberation or reasoning of his own: the expression of a commander's will that an act be done is intended to preclude or cut off any independent deliberation by the hearer of the merits pro and con of doing the act.³¹

He calls authoritative reasons that succeed in cutting off deliberation in this way "peremptory."

Raz, discussing the same topic, writes:

One thesis I am arguing for claims that authoritative reasons are pre-emptive: the fact that an authority requires performance of an action is a reason for its performance which is not to be added to all other relevant reasons when assessing what to do, but should exclude and take the place of some of them.³²

Following these thoughts, it might be objected that opacity and content inde-

The Morality of Freedom, chs. 2–3, and *Practical Reason and Norms*; and Rawls, "Two Concepts of Rules."

30 I thank an anonymous reviewer for encouraging me to address this point and for suggesting the term "opacity."

31 Hart, *Essays on Bentham*, 253.

32 Raz, *The Morality of Freedom*, 47 (emphasis removed from original).

pendence are analytically tied, and that the view I have proposed in this paper cannot accommodate this. For I have suggested that, in asking whether we have reasons to obey the law, we should look to the various distinctive normative properties of the law as grounds of such reasons, whereas opacity seems to require that we—in our role as subjects of the law—refrain from precisely this kind of inquiry.

There are several answers to Hart's and Raz's challenge. It is important to note, first, that content independence and opacity are not analytically tied *tout court*, in the sense that neither is a property of or entails the other, nor are they necessarily mutually coextensive.³³ That this is so is clear enough as a matter of reflection, I think, but can also easily be seen by considering the case of threats. Threats are archetypally content independent—one's reason to do as a threatener demands (if one has such reason) arises in virtue of the threat rather than in virtue of what it is demanded that one do—yet they are not also opaque. One thought, already mentioned, is that the grounds of opaque reasons are not transparent (or not meant to be transparent) to those to whom they apply. But of course, the transparency of the facts which ground a threat's normative force—viz., the badness of the threatened consequences together with the conditional assurance that they will be brought about—is crucial to the threat's functioning. Another thought, following Hart and Raz, is that opaque reasons replace or exclude from consideration other reasons bearing on the act in question. Yet threats do no such thing. Indeed, to weigh the reason given by the threat together with all of one's other reasons for and against doing the thing demanded seems *precisely* what is called for in such cases. Threats, then, are not plausibly sources of opaque reasons, though they are clearly content independent. So it cannot be that the two properties do not come apart.

It may be, however, that *in the domain of interest*, content independence and opacity always come together. This leads to a second point. For both Hart and Raz, opacity and content independence are discussed as part of their analyses of *authoritative reasons*. Moreover, while content independence is acknowledged to be a property of many reason types, the possession of the further property of opacity is what is said to be a distinctive characteristic of *authoritative* reasons. I shall consider in a moment whether anything I have said is incompatible with this conception of authoritative reasons. But before that, it is worth noting that

33 Neither Hart nor Raz appear to take opacity and content independence to be analytically linked in this sense. Hart discusses what he calls the “peremptory character” of authoritative reasons separately and before discussing their content independence. See Hart, *Essays on Bentham*, 254. Raz discusses content independence and “preemption” in different chapters entirely of *The Morality of Freedom*, and the two are nowhere discussed together in his major work on exclusionary reasons, *Practical Reason and Norms*.

I have not endeavored in this paper to offer an account of the authority of law, only to ask whether it provides content-independent moral reasons to do as it demands. Whether opacity is a feature of what it is for a reason to be authoritative, as Hart and Raz suggest, is not a matter we should expect our more modest question to pertain to.

This narrow point conceals a broader one. If the arguments in the remainder of this paper are correct, then we may conclude that the law often provides content-independent moral reasons to do as it demands, and moreover that these reasons often amount to obligations. This conclusion alone undermines a central strand of philosophical anarchism. The fact that we may reach this conclusion without asking the further question of whether the law, in providing these reasons, does so *authoritatively*—as well as what this means, and what it adds—is of theoretical interest in itself. We should take care to distinguish, in thinking about the normativity of the law, the question of whether it has reason-giving moral force from the question of whether it is authoritative. It is a virtue of the approach taken here that it allows us to see the space between these two questions.³⁴

The challenge that remains is that the view proposed in this paper is incompatible with the opacity of authoritative reasons to obey the law. If Hart or Raz is correct about the character of the authority of the law, and if we do indeed have authoritative reasons to obey the law, then this incompatibility would count as a strike against this paper's main claims. I believe, however, that the view of content independence proposed here is compatible with Hart's and Raz's notions of opacity. To see this, it will help to describe more precisely their views.

For Raz, an authoritative reason is really two reasons: a "first order" reason for action as well as an "exclusionary reason," which is a species of "second order" reason:

A second order reason is any reason to act for a reason or to refrain from acting for a reason. An exclusionary reason is a second order reason to refrain from acting for some reason.³⁵

An authoritative reason to ϕ is thus, on this view, both a first-order reason to ϕ

34 The topic of the law's authority is too large and difficult a topic to enter into in this already lengthy discussion. I take up the question of the law's authority in further work. See Flanagan, "Essays at the Limits of the Law."

35 Raz, "Reasons for Action, Decisions and Norms," 487. The nature of second-order reasons is a point of controversy. See Piller, "Kinds of Practical Reasons"; Whiting, "Against Second-Order Reasons"; and Scanlon, "Reason and Value." These are recent discussions, though as Raz notes in the postscript to *Practical Reason and Norms*, the history of this debate goes back at least four decades.

as well as a second-order reason to refrain from acting for other reasons favoring or disfavoring ϕ -ing. This structure is familiar, Raz claims, from promises, which also generate exclusionary reasons. By way of illustration (to borrow one of Raz's examples), if one has promised to consider only one's child's interests in decisions about the child's schooling, then one has an exclusionary reason not to make such decisions for (among other things) self-interested reasons, such as that it would require a career sacrifice, mean fewer vacations, require substantial time driving to and from the campus, and so on. These reasons are not to be considered in deliberation about how to best educate the child.

Though he does not say so in exactly these terms, what Hart has in mind is something similar. Here he is again:

The expression of a commander's will that an act be done is intended to preclude or cut off any independent deliberation by the hearer of the merits pro and con of doing the act.³⁶

Insofar as Hart is making a claim about the character of authoritative reasons (rather than the mental states of those issuing commands), we should read him as claiming that, in addition to being a reason to ϕ , an authoritative reason also has the effect of prohibiting further deliberation about whether to ϕ . A strong reading of this prohibition would preclude any contemplation of the merits of ϕ -ing; a weak reading would only preclude such considerations from figuring in deliberation about whether to ϕ .³⁷ On either reading, Hart is suggesting, like Raz, that authoritative reasons also regulate a subject's proper consideration of the various other reasons that may favor or disfavor ϕ -ing, and so we may treat both as holding the view that authoritative reasons entail exclusionary reasons.

Hart's and Raz's claims are thus about the scope of reasons *for which* it is appropriate for subjects to ϕ when the law authoritatively demands that they ϕ , as a matter of practical deliberation. It is *not* about the considerations that *in fact* count in favor of ϕ -ing, nor is it about the further facts that ground the law's authority in that case. It might therefore be the case both that (as I have claimed)

1. [The law has some properties $\{P\}$] and [The law demands that $S \phi$] together ground [S has a reason to ϕ],

and that (as Hart and Raz claim)

2. The reason *for which* S should ϕ is [The law demands that $S \phi$] together with [The law is authoritative].

³⁶ Hart, *Essays on Bentham*, 253.

³⁷ Raz attributes the strong reading to Hart. See Raz, *The Morality of Freedom*, 39.

There is no incompatibility here: 1 is about the grounds of subjects' reasons to obey the law, and 2 is about the reasons for which subjects should act when they consider whether to do as the law demands. If it is *also* true (though it might not be) that the properties of the law $\{P\}$ are those that make the law authoritative, then 1 will help explain 2. (Subjects must also deliberate regarding whether a reason is authoritative, in order to know whether 2 applies to them; 1 may in that case help them to do so.) If not, then the reasons grounded by those properties will be superfluous to a subject's deliberation whether to ϕ . In either case, if these theories about authority are correct, there is a deliberative opacity between 1 and 2 but no conflict. Again, it is a virtue of the approach taken here that it allows us to clearly see this relationship.

As I said, there is another version of this objection. This objection is deeper, because according to it, content independence and opacity emerge as structural or logical consequences of the law as a rule-governed practice, such that moral reasons to obey the law because it is the law *must* be opaque.³⁸ A version of this objection is suggested by Rawls, who argues that "the rules of practices are logically prior to particular cases," which is to say that

given any rule which specifies a form of action (a move), a particular action which would be taken as falling under this rule given that there is the practice would not be *described as* that sort of action unless there was the practice. In the case of actions specified by practices it is logically impossible to perform them outside the stage-setting provided by those practices, for unless there is the practice, and unless the requisite proprieties are fulfilled, whatever one does, whatever movements one makes, will fail to count as a form of action which the practice specifies. What one does will be described in some *other way*.³⁹

As an example, Rawls offers the rules of baseball. In baseball, to record three strikes at bat *just is* to strike out, and the act of "recording a strike," as well as the states of being "at bat," "out," and so on are *defined by* the rules of baseball. Thus, the question

Shall I stay at bat after my third strike?

can *only* be answered, and an answer can *only* be justified, by reference to the rules of baseball. After a third strike, one has struck out; and once one has struck

38 Note that this objection does not claim opacity is a property of content independence generally, just that the two are coextensive in the domain of reasons to obey the law, and other structurally similar domains, in virtue of the structure of those domains.

39 Rawls, "Two Concepts of Rules," 25 (emphasis in original).

out, one is no longer to be at bat. This is true even though in another frame of mind, we might wonder whether the rules of baseball are the best rules, should be amended, and so on. As a player, Rawls argues, such considerations simply do not bear on the question of what to do, and so are excluded from a player's deliberation about how to act.

Since this question is about one of the many rules of baseball, the same might be thought true of the general question

Shall I obey the rules of baseball?

Likewise, the answer to this question might be thought to be given only by the fact of whether one is playing baseball or not. If one is playing, the rules simply apply to one, and further considerations are excluded. So while one might ask in the first instance *whether* to play baseball or not, once one has decided to play, there is no further question of whether to play as the rules demand. To play just is to have the rules apply. The general question is thus necessarily answered in the affirmative.

This way of conceiving rules of practice renders the reasons given by them content independent. An individual player is to take as her reason for (say) returning to the dugout the fact that she is out, which is made true by the three strikes she has had at bat. Both are made true by the rules of baseball, which might have defined being "out" differently or which might have required something else of a player who strikes out. Her reason to return to the dugout is thus grounded by the fact that the rules apply to her together with the fact of what the rules require, rather than by the independent merits of returning to the dugout or not. It is, in this way, content independent.

Importantly, this form of content independence is also *opaque* with respect to the underlying justification of the rules. This is because the normative properties of the rules which justify *them*—e.g., that they are most conducive to fun, competition, fairness, and so on—do not also ground the fact that the player is out *or* the fact that the rules apply to her.⁴⁰ Those properties do not therefore ground the player's reason to return to the dugout. Since the normative properties of the rules do not ground the player's reason to obey them, the reason is therefore opaque with respect to those properties, and they are excluded as reasons bearing on what she is to do.

40 This is slightly too quick. The same considerations that justify the rules may, depending on our view, also play some part in grounding the fact that they apply to players. Even if they do, however, variations in those considerations—e.g., just *how* conducive to fun this rule is—would affect a player's reason to obey only if they altered the fact of whether the rules applied or not, which presumably most such considerations would not. The player's reason to obey the rule, then, would remain opaque with respect to those considerations.

It is easy to see, by analogy, how the same might be thought true of the law and our reasons to do as it demands. For just as some acts and states are defined by the practice of baseball, so too we might think that some acts and states are defined by the law.⁴¹ Thus, perhaps the question

Shall I pay x dollars to the tax collector?

can only be answered by reference to the law of the land, since the tax regime, the office of the tax collector, and even the dollar and the very notion of *tax* are defined by the law. Again, it might be thought that the tax law simply *applies* to one, such that there is no intelligible question of whether to obey. (Or at least that this is true *insofar as* the tax law applies to one.) And while there may be good reasons to adopt *this* tax code or *that* one, such questions apply at the level of legislation and regulation, not to an individual subject wondering whether to write a check for x dollars to the revenue service.

Again, likewise, it might be thought that this question about taxes is just one instance of the more general question

Shall I obey the law?

and that the interrogative form of this general question disguises the important truth that the law *simply applies* to those subject to it. We are all playing, and the rules apply to us all.

If this were true, then content independence would similarly be a structural property of the way the law creates demands on its subjects, and the reasons for those subjects to obey would be opaque with respect to the considerations that justify the law itself.

However, whatever conclusions we should draw from Rawls's argument, it should be clear that they cannot be these. We can intelligibly ask whether or not to obey the law, or *this law* or *that law*, even when the acts we are considering whether or not to do are *defined by* the law; and the mere fact of the law's application to us cannot settle that question. Indeed, we can see on reflection that precisely the same is true of a player wondering whether to obey the rules of a game.

This is, first, because on Rawls's analysis, a question of the form

Shall I stay at bat after my third strike?

and in the relevantly similar sense, so too the question

⁴¹ Rawls makes this point by offering the example of rules of punishment. I think the example of taxes is an easier one, and so I shall proceed with it, but nothing is meant to hang on this choice. See Rawls, "Two Concepts of Rules," 10–18.

Shall I pay x dollars to the tax collector?

is a question not about *what to do* but rather a question about *what the rules are*. They are thus not particular instances of more general questions about whether to obey the rules. To the contrary, that general question can still be asked in both cases.⁴² Perhaps the player has reasons to break the rules by staying at bat after a third strike—reasons from within the game, such as that it would beneficially delay the game, or reasons from without, such as that doing so would serve as some political protest.⁴³ Then the player may sensibly ask whether or not to obey that rule, as a way of asking whether to continue within the practice. Likewise, those subject to the law may sensibly ask whether or not to do as it demands. We can ask this question from within the practice, and considerations from both within and outside of the practice can bear on the answer. Since we are concerned here with the question of whether to obey the law, not the question of what the law demands when it applies to us, it is important that the former question is seen *not* to be opaque with regard to reasons that come from outside the practice. It is, rather, transparent in precisely the way I have been suggesting so far.

Second, Rawls himself suggests that questions about acts that are practice defined may be redescribed in other, nonpractice-defined terms, or may inevitably entail giving answers to such questions.⁴⁴ Thus to answer the question

Shall I stay at bat after my third strike?

might also be described as, or entail an answer to, the question

Shall I continue standing here, on this spot, now?

The latter question makes no necessary reference to practice-defined states or acts, and so may be answered by reference to all of the reasons for and against continuing to stand in that spot. (Perhaps you promised you would do so; perhaps you are being threatened not to move; etc.) Importantly, the answer to this

42 Rawls recognizes this point when he writes that if one's appeal to the rules is not accepted, "it's a sign that a different question is being raised as to whether one is justified in accepting the practice, or in tolerating it." See Rawls, "Two Concepts of Rules," 27.

43 Whether strategic rule breaking is an act within a practice or an abrogation of it is a vexed question that I cannot enter into here. On this question with respect to Rawls and baseball, see Palmiter, "Cheating, Gamesmanship, and the Concept of a Practice." There is also a lively debate regarding this question with respect to contracts and promises. See, e.g., Shiffrin, "The Divergence of Contract and Promise."

44 Rawls again: a practice-dependent act "would not be *described as* that sort of action unless there was the practice. . . . What one does will be described in some *other way*" ("Two Concepts of Rules," 25, emphasis in original).

wider question may *supersede* the answer to the first, since it includes in it the value of abiding by the rules of the practice versus the disvalue of breaking them.

The point is easier to see when we consider certain questions of law. For Justice Lemuel Shaw, to answer the question

Shall I order Thomas Sims returned to slavery in accordance with the Fugitive Slave Act?

was also to answer the question

Shall I bring it about that Thomas Sims is re-enslaved?

The first question might have been decidable only by the standards of the law, but to answer that question was at the same time to answer the second question, which clearly involves wider moral considerations, including those concerning the value of the law itself. (This is true, I think, even if the state of *enslavement* is taken to be defined by a legal practice.) To argue that Shaw's question was fully settled by the law, or that the law simply applied to Shaw's judicial act, is to ignore the clear conflict between the demands of the statute and the requirements of morality.⁴⁵

We can put the point generally. Systems of rules generate, for those subject to them, content-independent reasons: insofar as one is aiming to act within a practice, one's reason to ϕ arises in virtue of the rules of the practice, rather than in virtue of facts about ϕ -ing as such. From within the practice, one's reason to ϕ may be opaque. But one can also ask whether the act required by the practice can be redescribed in other terms, or whether to decide to do that act is also to decide to do some further act. Wider considerations may bear on this further question, and the answer to it may impinge on the answer to the first. The opacity of rule-given reasons, then, may in this way be made transparent.

If I am right about these ways to answer Rawls's challenge, then this version of the objection fails as well. Even for practice-defined acts demanded by the law, we can ask whether we ought to do them. This question is not opaque: it calls for us to think about the various normative properties of the law that might give us reason to obey or disobey. It also, in the manner I have been describing, remains content independent: if we have reason to obey, it will be in virtue of those distinctive properties of the law, rather than in virtue of properties of the thing the law demands we do.

45 Shaw infamously ordered Sims returned to slavery. For details of the case, see Brown, "Thomas Sims's Case after 150 Years."

4. ANARCHISM

Even if it may be claimed, as I have argued it may, that there *could be* genuinely normative, content-independent moral reasons for action given by the law, it remains to be shown that such reasons do in fact exist (or, as I shall more modestly claim, that it is plausible that such reasons exist). We can turn finally, then, to the question of whether, and in what sense, we might truly claim about the law that it provides such reasons.

Let us begin with the easier issue of *in what sense* we might truly make such claims. As I noted at the beginning of this essay, it is telling that the “traditional question” in the study of legal obligation, and indeed the name of the field itself, concerns not *reasons* to obey the law but our *obligation* to do as it demands. We may restate this traditional question of whether we are *obligated* to obey the law, because it is the law, as the “obligation question”:

Does the law provide genuinely normative, content-independent, and normally decisive moral reasons to do as it demands?⁴⁶

In order to answer yes to this question, we would need to identify some property or properties of the law which, together with the law’s demanding that we act, fully ground our having normally decisive moral reasons to so act.

I do not doubt that historically, some have thought it plausible to provide an affirmative answer to the obligation question and to identify such properties. We might, for instance, agree with the First Vatican Council that

1. the pope is the earthly representative of God and is preserved from the possibility of error,
2. the law as handed down by the pope (the “pope’s law”) is normatively binding, in the sense that we each have decisive reason to do as it demands,

and that therefore

3. the pope’s law provides each of us with genuinely normative, content-independent, and decisive moral reason to do as it demands.⁴⁷

Alternatively, as in an earlier example, we might hold a similar view about the obligation of subjects to obey the laws handed down to them by monarchs, in

46 As explained in section 1, I use “normally decisive reason” as a moderately ecumenical analysis of the concept of obligation.

47 Here, as before, it is helpful if we understand the pope’s law as something more than a mere codification of God’s law.

view of the natural right of kings and queens. Such a right might, like the pope's, or as on Filmer's view, be grounded in divine right or revelation, or else, as on the common reading of Hobbes, be grounded in the necessary covenant of each to every other in order to gain protection from the war of all against all.

Such defenses of the affirmative answer to the obligation question do not, however, strike me as plausible, nor do I suppose they strike many as plausible today. And yet, despite this, writers persist in treating the obligation question as the one that demands an answer.⁴⁸ This insistence, unsurprisingly, has led many to answer no, and instead to endorse some version of philosophical anarchism. For if we must endorse either the view that the law always gives us genuinely normative, content-independent, and normally decisive moral reasons to do as it demands, or the view that it gives us no reasons at all, the anarchist's choice is clearly the best one. Faced with such a dilemma, it would be difficult to adopt any other position.

But this dilemma, I think it is clear, is a false one. We should not normally expect the law to in all cases give us decisive reason to do as it demands.⁴⁹ Rather, I believe we should expect the law in many cases to *add to the balance of reasons* in favor of doing as it demands, by providing *some reason* for action. The strength of the reasons so provided by the law may vary according to which property or properties give it normative force, but the reasons should be perceptible nonetheless when we look for them. Often, such reasons will do the more important job not merely of providing some reason to act but of contributing, alongside other reasons, to making it the case that we ought to do as the law demands, in the decisive reason-implicating sense—and thus in part, we can add, to contributing to making it the case that we ought to do the thing the law demands *because the law demands it*. And occasionally, or so I shall argue, legal reasons may ground our obligations to do as the law demands not merely *in part* but rather *fully make it the case* that we ought to do as the law demands. In this way, then, we can see the law as giving us genuinely normative, content-independent moral reasons to do as it demands, in a way that does not amount or tend to any version of philosophical anarchism.

We can now turn to the more difficult question,

Which property or properties of the law, together with its demanding that we act, may plausibly give us a reason to so act?

48 See *inter alia* Dagger and Lefkowitz, "Political Obligation"; Green, "Legal Obligation and Authority."

49 I think we should make this claim even if we believe, as some do, that the law claims for itself the authority to create obligations, in the decisive reason-implicating sense.

It is worth emphasizing again that I shall pursue only the modest goal of attempting to show that certain properties of the law may *plausibly* help give us genuinely normative, content-independent moral reasons for action, rather than attempting to mount a conclusive or even very strong argument that they do so. I shall do this by briefly sketching the ways in which theories of political and legal *obligation* may be easily and plausibly recast as theories of political and legal *reasons*. And although this weakens these theories, it also makes them more plausibly *true*; and moreover, when viewed in this way, I think we may more clearly see how these theories may *together* give us something resembling a blanket obligation to obey the law.⁵⁰

I should first point out that many of the leading nonvoluntarist theories of political and legal obligation may be recast as more modest theories of sources of political and legal reasons. Fair-play theories provide a clear example. When Nozick objects, for instance, that we cannot come under an obligation to others simply because they have conferred some benefit upon us, we may answer that the conferral of certain benefits may nonetheless generate *some reason* for us to participate appropriately in the system of benefits.⁵¹ Such an answer is plausible even in his famous public address system case: when it is your turn, you are, let us agree, not obligated to perform, in the sense that you do not have a decisive reason to do so, but if you have enjoyed the fruits of the cooperative enterprise, then you may plausibly have *some reason* of fairness to do your part in the future. In the analog case of the state, this may result in reasons of fairness to obey the law. (I do not argue that this is clearly true, only that it is very plausibly true—and much more plausibly true than its original obligatory counterpart.)

Or, to take a similar example, when Klosko claims that

1. if some state is a cooperative enterprise, and
2. if this state, through its laws, provides its citizens with presumptively beneficial, fairly distributed goods,

he might conclude either that

3. the state's citizens have an obligation of fairness to obey its laws,

or, instead, more modestly that

4. the state's citizens have reasons of fairness to do as its laws demand.

Klosko's conclusion is in fact 3 but it need not be: 4 is a weaker conclusion and is thus easier to establish and open to fewer objections. It is also, I think, much

50 See Klosko, "Multiple Principles of Political Obligation," who explores a version of this view.

51 Nozick, *Anarchy, State, and Utopia*, 90–95.

more immediately plausibly true. The reasons given by 4 may or may not amount to an obligation to do as the law demands, but would rather provide us with *some reasons*, by counting in favor of our doing so.

Note here one important fact, which is that for theories like this to successfully give us genuinely normative content-independent *legal* (and moral) reasons for action, it must be some part of *the law* that provides us with the goods whose receipt grounds our reasons to obey the law's demands. That is to say, if we have reasons in virtue of some principle of fairness to do our part in some collective enterprise, those reasons are only *legal reasons* of the kind we have been here discussing if the product of the collective enterprise is in some way secured *by the law*. Otherwise, our reasons to do our part will be just that, and any specifically legal demand that we do so will merely restate those reasons rather than giving us new, additional ones.

Other nonvoluntarist theories of political and legal obligation may be similarly recast in this way, including theories built around principles of gratitude, samaritanism, and natural duty. That is because all such theories identify some moral principle that, they argue, is operative in virtue of the existence or some other feature of the law. Any such theory, as in the case of fair-play theories, may more easily establish that the moral principles they identify provide some reason to obey the law than that they provide normally decisive reasons to do so.

Recasting these theories in this way has benefits beyond making their conclusions easier to establish. Understood as independent sources of reasons to obey the law, these theories may be very naturally combined to generate stronger reasons to obey the law than any one of them provides on its own. It may also be the case that some principles provide reasons to obey the law in only some rather than all domains, or reasons whose strength varies across different domains of the law. Combining such principles may allow us to claim that there are widespread reasons to obey the law, because it is the law, in ways that would be impossible drawing on any one principle alone. Acknowledging this possibility, moreover, may help our theories better match our sense that it is in some cases much more important to obey the law than in other cases. The question facing these recast theories is thus not, "Is the principle relied upon by this theory sufficient to generate wide-ranging obligations to do as the law demands?" but, much more modestly, "Does the principle relied upon by this theory generate reasons to do as the law demands?"

A similar point may be made about theories attempting to ground an obligation to obey the law in the fact of some laws' democratic provenance. Such theories claim that

[L was generated by a democratic process] and [L demands that $S \phi$] together ground [S has an obligation to ϕ].

Naturally, however, we might instead make the more modest claim that

[L was generated by a democratic process] and [L demands that $S \phi$] together ground [S has a reason to ϕ].

In these claims, of course, [L was generated by a democratic process] stands in for a more complex statement of the feature or features had by democratic laws in virtue of which those subject to them may be obligated or have reason to do as they demand. When Christiano argues, for instance, that (roughly) democratically created laws treat each citizen equally with respect to certain questions about what we together should do, regarding which no one has greater claim or standing to give an answer than any other, or when Kolodny writes that “the concern for democracy is rooted in a concern not to have anyone else above—or, for that matter, below—one,” each is arguing that it is *this* more specific feature of democratically created laws which in part grounds our obligation to obey those laws.⁵²

This is not, I should say again, the place to engage in a discussion of whether Christiano’s or Kolodny’s claims, or those of other democratic theorists, about democratic political obligation succeed, nor do I here mean to endorse either’s claim to that effect or my suggested weaker version of those claims. Rather, I mean only to claim that these arguments provide plausible accounts of one source of genuinely normative, content-independent moral reasons to do as the law demands; and that, as with the nonvoluntarist accounts I made similar claims about earlier, such accounts of the source of legal reasons may be combined with others, and may vary in presence and strength across different domains of the law.

Some consent theories may also be recast in this way. According to consent theories, S is obligated to obey L just when and because some combination of 1–4 holds:

1. *Ordinary consent*: S has consent to do as L demands.
2. *Tacit consent*: S has tacitly consented to do as L demands.
3. *Hypothetical consent*: S would so consent if S knew all the facts, deliberated rationally, and so on.
4. *Normative consent*: S should so consent.

Of course, if S has consented to obey the law, it may often be accurate to claim

52 Christiano, “The Authority of Democracy”; Kolodny, “Rule over None II.”

that S is obligated to do as it demands, because S's consent grounds a normally decisive reason to do so. But it is worth emphasizing with respect to conditional and normative consent theories that these theories may be much more plausible as theories of the sources of *reasons* to do as the law demands rather than the sources of *obligations* to do so. It is much more plausible, for example, to argue that [S would consent to obey the law] gives S a *reason* to obey the law than that it gives her an *obligation* to do so.⁵³ (Of course, such a reason may *alongside others* amount to a decisive reason. In this way, conditional or normative consent to a law may partially ground an obligation to obey it.)

There is one other plausible source of genuinely normative, content-independent moral reasons to obey the law that I wish to take a bit more time over here, because I believe it may be of particular importance. This source is the law's often unique ability to solve coordination problems.

It will help to consider the case of traffic laws. Those of us who drive each have some reasons *not* given by the law to drive in certain ways: these can be helpfully summed up by saying that we have all the summary reason to *drive safely*. One of the reasons summarized by this reason is the reason we all have to drive on one side of the road; another is to drive at a safe speed. But these reasons are in an important sense incomplete. If we are driving, say, on many highways in the United States, we have reasons to drive on the *right* side of the road and to drive in the vicinity of 55 mph. When we are in other places and on other roads, these reasons change. But wherever we are, these reasons are grounded by facts about the law.

This last claim might be doubted. The speed at which we have reason to drive on some road, for instance, is determined partly by the road itself, by the capabilities of our cars, and by how fast and how many others are traveling. It may seem that the legal speed limit is superfluous, or that it merely formalizes these other reasons. But this argument neglects the further reason we all have to drive in the vicinity of some single, particular speed. Which speed this is may be *limited* by the road, our cars, and how many of us there are, but this speed is not *fully determined* by these facts. The law accomplishes this latter task.

Similarly, it may be rightly pointed out that the law is not a necessary ground of our reasons to drive on this or that side of the road. If there were no law concerning which side of the road to drive on, people might just work out for themselves some convention. If they did, these people would have a reason to drive on whichever side of the road that convention dictated.

Equally, if the law in some place demanded that we all drive on the left,

53 See Enoch, "Hypothetical Consent and the Value(s) of Autonomy"; Stark, "Hypothetical Consent and Justification."

whereas in fact everyone followed the practice of driving on the right, each driver would have most reason to drive on the right—and it is plausible to think that each driver would have *no reason* to drive on the left.

But this is not an objection to the view I am defending. I do not claim that the law is a necessary ground for our reasons to drive on one side or another, or to drive at a certain speed; I am claiming only that it is *in fact* the ground of many of our actual such reasons. It is not enough to say that we each have a reason to drive on the side of the road on which most other drivers drive. *We together* must at some point take some actions or decisions which determine the particular side that is. This could take the form of legislation, or it could be established through more complex patterns of convention. In the actual case of the United States, I submit that it is the law that secures the relevant convention; it is the fact that the law demands that we drive on the right which partially grounds our reason to drive on the right. In other words, we cannot state the facts which ground

[we each have reason to drive on the right side of the road in the United States]

or

[we each have reason to drive near 55 mph on certain highways in the United States]

without making reference to the fact that the law demands that we do so. We may thus, as I argued earlier, call our reasons to drive in these ways *content-independent legal reasons*.⁵⁴

Reasons given by coordination problems solved by the law such as these may be, I believe, quite weighty reasons. Very seldom will I have sufficient reason to drive on the side of the road other than the side demanded by the law. I think it is therefore fair to say that we are *obligated* to drive on the side of the road demanded by the law, and we are so obligated *because it is what the law demands*.

We can next observe that traffic laws are not a special case, but rather one of very many sets of laws whose purpose is to solve coordination problems. I shall not defend this claim at length here, except to mention that many of the core functions of political organization are to help us live our lives together, and include the establishment of property regimes, monetary systems, rules of ex-

54 It might be further objected that the law here merely *happens* to provide the relevant convention—that the reasons to drive on the right side of the road, or around 55 mph, are not grounded by the law *in virtue of its being the law*, or in other words by the law's authority *as such*, but rather by the law *in virtue of its establishing the relevant convention*. But the fact that the law establishes certain conventions may quite plausibly be part of what grounds its authority in the relevant domains.

change, and indeed traffic laws, all of which are at least partly conventional; and so the reasons we may have with respect to these domains of law will be at least partly grounded in the fact of the law's demands.⁵⁵ If I am right that conventionally determined reasons of this kind are genuinely normative, content-independent moral reasons for action, then it seems that they are quite widespread and quite forceful. On their own they might license my claim that we do plausibly often have genuinely normative, content-independent moral reasons to obey the law; and combined with the other plausible sources of such reasons I have already mentioned, we may well be obligated to obey the law, because it is the law, much more often than we might otherwise have thought.

5. CONCLUSION

It may now help to sum up some of my main claims. I have argued that when we consider the question,

Do we have reasons to do as the law demands, because it is what the law demands?

we should understand the "because" in the question as the because of grounding. On the view I have defended, if we have such reasons, it is because there is some fact about the law that at least partially grounds the fact that we have such a reason. This, on my view, is what we should mean when we claim that the law may be a source of content-independent reasons to do as it demands.

I argued next that, once we see that this is what it is to be a content-independent reason to obey the law, we can see that we very plausibly have many such reasons. This is because many of the leading theories of political and legal obligation may be recast as theories about content-independent reasons to do as the law demands. When recast in this way, these theories' main claims are easier to establish; and although they are thereby individually weaker, they may gain

55 David Lewis mentions this in *Convention*, and there is also a sizeable jurisprudential literature concerning coordination, convention, and the law. See *inter alia* Gans, "The Normativity of Law and Its Co-Ordinative Function"; Ullmann-Margalit, "Is Law a Co-Ordinative Authority?"; Postema, "Coordination and Convention at the Foundations of Law"; Green, "Law, Co-Ordination and the Common Good"; Green, *The Authority of the State*, ch. 4; and Raz, *The Morality of Freedom*, 30. Marmor admits that coordination problems can ground obligations but also claims that this cannot explain the full extent of the law's normative power ("The Dilemma of Authority"). Like Ripstein, I am inclined to disagree. On Ripstein's construction of Kant's political philosophy, nearly all of our political duties are "determined" in this way by the state and the state's laws. For some considerations along these lines, see Ripstein, *Force and Freedom*, as well as Pallikkathayil, "Deriving Morality from Politics"; Julius, "Independent People" and "Public Transit."

the advantage of being *true*, and they may also be combined with each other to provide stronger summary reasons to obey the law. Understanding our reasons to obey the law in this way, I claimed, may also help explain our intuition that such reasons may vary in strength across various circumstances and the various domains of the law. If these theories may be recast and combined in this way, I also argued, and if the law is a source of content-independent reasons for action in the many cases in which it helps solve coordination problems, then such legal reasons may very often be sufficiently strong to make it the case that we *ought* to obey the law. In other words, we may often have more than content-independent moral *reasons* to obey the law—we may have *obligations* to do so.

I have also argued that we need not be philosophical anarchists just because we believe that no one theory of political and legal obligation has successfully established such an obligation. We should, I suggested, be engaged in the more modest enterprise of looking for reasons to obey the law, and then investigating their strength and the domains over which they range. In this way, I believe, we are likely to find a picture of our reasons for obeying the law that more accurately reflects our considered views, and, importantly, a picture that does not tend toward anarchism.⁵⁶

Harvard University
eflanigan@g.harvard.edu

REFERENCES

- Berker, Selim. "The Unity of Grounding." *Mind* 127, no. 507 (July 2018): 729–77.
- Brown, Thomas J. "Thomas Sims's Case after 150 Years: A Motion to Reconsider." *Massachusetts Legal History: A Journal of the Supreme Judicial Court Historical Society* 7 (2001): 95–118.
- Christiano, Thomas. "The Authority of Democracy." *Journal of Political Philosophy* 12, no. 3 (September 2004): 266–90.
- Cohen, G. A. *Rescuing Justice and Equality*. Cambridge, MA: Harvard University Press, 2009.
- Dagger, Richard, and David Lefkowitz. "Political Obligation." *The Stanford Ency-*

⁵⁶ I owe special thanks to Selim Berker and Derek Parfit for their extensive and extremely helpful comments on an early draft of this paper. Small tribute though it is, I dedicate this paper to Derek Parfit *in memoriam*. I am also particularly indebted to James Brandt, who read and commented on several drafts. For further helpful comments and discussion, I thank Adriana Alfaro Altamirano, Arthur Applbaum, Eric Beerbohm, Noam Gur, Chris Havasy, and two anonymous reviewers for this journal.

- lopedia of Philosophy* (Fall 2014). <https://plato.stanford.edu/entries/political-obligation>.
- Edmundson, William A. "State of the Art: The Duty to Obey the Law." *Legal Theory* 10, no. 4 (December 2004): 215–59.
- Enoch, David. "Hypothetical Consent and the Value(s) of Autonomy." *Ethics* 128, no. 1 (October 2017): 6–36.
- . "Reason-Giving and the Law." In *Oxford Studies in Philosophy of Law*, vol. 1, edited by Leslie Green and Brian Leiter, 1–38. Oxford: Oxford University Press, 2011.
- Estlund, David. *Democratic Authority*. Princeton, NJ: Princeton University Press, 2008.
- Fine, Kit. "Guide to Ground." In *Metaphysical Grounding*, edited by Fabrice Correia and Benjamin Schnieder, 37–80. Cambridge: Cambridge University Press, 2012.
- Flanigan, Edmund. "Essays at the Limits of the Law." PhD diss., Harvard University, forthcoming.
- Gans, Chaim. "The Normativity of Law and Its Co-Ordinative Function." *Israel Law Review* 16, no. 3 (1981): 333–49.
- Green, Leslie. *The Authority of the State*. Oxford: Clarendon Press, 1988.
- . "Law, Co-Ordination and the Common Good." *Oxford Journal of Legal Studies* 3, no. 3 (Winter 1983): 299–324.
- . "Legal Obligation and Authority." *The Stanford Encyclopedia of Philosophy* (Winter 2012). <https://plato.stanford.edu/entries/legal-obligation>.
- Gur, Noam. "Are Legal Rules Content-Independent Reasons?" *Problema* 5 (2011): 175–210.
- Hart, H. L. A. *Essays on Bentham: Jurisprudence and Political Philosophy*. Oxford: Oxford University Press, 1982.
- . "Legal and Moral Obligation." In *Essays in Moral Philosophy*, edited by A. I. Melden, 82–107. Seattle: University of Washington Press, 1958.
- Jay-Z and Kanye West. "No Church in the Wild." Track 1 on *Watch the Throne*. Roc-A-Fella Records, Roc Nation, and Def Jam Recordings, 2011.
- Julius, A. J. "Independent People." In *Freedom and Force: Essays on Kant's Legal Philosophy*, edited by Sari Kisilevsky and Martin Stone, 91–110. London: Hart Publishing, 2017.
- . "Public Transit." In *Reasons and Intentions in Law and Practical Agency*, edited by George Pavlakos and Veronica Rodriguez-Blanco, 159–75. Cambridge: Cambridge University Press, 2015.
- Klosko, George. "Are Political Obligations Content Independent?" *Political Theory* 39, no. 4 (2011): 498–523.

- . “Multiple Principles of Political Obligation.” *Political Theory* 32, no. 6 (2004): 801–24.
- Kolodny, Niko. “Rule over None II: Social Equality and the Justification of Democracy.” *Philosophy and Public Affairs* 42, no. 4 (Fall 2014): 287–336.
- Lewis, David. *Convention*. Cambridge, MA: Harvard University Press, 1969.
- Locke, John. *Two Treatises of Government*. Edited by Peter Laslett. Cambridge: Cambridge University Press, 1988.
- Markwick, Paul. “Independent of Content.” *Legal Theory* 9, no. 1 (March 2003): 43–61.
- . “Law and Content-Independent Reasons.” *Oxford Journal of Legal Studies* 20, no. 4 (Winter 2000): 579–96.
- Marmor, Andrei. “The Dilemma of Authority.” *Jurisprudence* 2, no. 1 (2011): 121–41.
- Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.
- Pallikkathayil, Japa. “Deriving Morality from Politics: Rethinking the Formula of Humanity.” *Ethics* 121, no. 1 (October 2010): 116–47.
- Palmiter, Brian. “Cheating, Gamesmanship, and the Concept of a Practice.” Unpublished manuscript.
- Parfit, Derek. *On What Matters*. 3 vols. Oxford: Oxford University Press, 2011–2017.
- Piller, Christian. “Kinds of Practical Reasons: Attitude-Related Reasons and Exclusionary Reasons.” In *Analyses*, edited by S. Miguens, J.A. Pinto, and C. E. Mauro, 98–105. Porto: Porto University, 2006.
- Postema, Gerald J. “Coordination and Convention at the Foundations of Law.” *The Journal of Legal Studies* 11, no. 1 (January 1982): 165–203.
- Rawls, John. *A Theory of Justice*. Rev. ed. Cambridge, MA: Harvard University Press, 1999.
- . “Two Concepts of Rules.” *The Philosophical Review* 64, no. 1 (January 1955): 3–32.
- Raz, Joseph. *The Morality of Freedom*. Oxford: Clarendon Press, 1986.
- . *Practical Reason and Norms*. 2nd ed. Oxford: Oxford University Press, 1999.
- . “Reasons for Action, Decisions and Norms.” *Mind* 84, no. 336 (October 1975): 481–99.
- Ripstein, Arthur. *Force and Freedom*. Cambridge, MA: Harvard University Press, 2009.
- Rosen, Gideon. “Metaphysical Dependence: Grounding and Reduction.” In *Modality*, edited by Aviv Hoffmann and Bob Hale, 109–36. Oxford: Oxford University Press, 2010.

- Ross, W.D. *The Right and the Good*. Edited by Philip Stratton-Lake. Oxford: Clarendon Press, 2002.
- Scanlon, T.M. *Being Realistic about Reasons*. Oxford: Oxford University Press, 2014.
- . “Reasons: A Puzzling Duality?” In *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, edited by R.J. Wallace, Philip Pettit, Samuel Scheffler, and Michael Smith, 231–46. Oxford: Clarendon Press, 2004.
- Sciaraffa, Stefan. “On Content-Independent Reasons: It’s Not in the Name.” *Law and Philosophy* 28, no. 3 (May 2009): 233–60.
- Shiffrin, Seana Valentine. “The Divergence of Contract and Promise.” *Harvard Law Review* 120 (2007): 708–53.
- Smith, Matthew Noah. “Political Obligation and the Self.” *Philosophy and Phenomenological Research* 86, no. 2 (March 2013): 347–75.
- Stark, Cynthia A. “Hypothetical Consent and Justification.” *The Journal of Philosophy* 97, no. 6 (June 2000): 313–34.
- Ullmann-Margalit, Edna. “Is Law a Co-Ordinative Authority?” *Israel Law Review* 16, no. 3 (1981): 350–55.
- Whiting, Daniel. “Against Second-Order Reasons.” *Noûs* 51, no. 2 (June 2017): 398–420.
- Wolff, Robert Paul. *In Defense of Anarchism*. New York: Harper and Row, 1970.

THE AMBITIONS OF CONSEQUENTIALISM

Brian McElwee

A CONSEQUENTIALIST ACCOUNT of some subject gives an assessment of it by reference to the production of good outcomes. This characterization leaves open (at least) four issues:

1. What is the subject being evaluated, e.g., acts, rules, motives, character traits?
2. Which components of an outcome contribute to making that outcome good, e.g., the promotion of well-being, the promotion of the glory of god, the universe being well ordered?
3. What distribution of those components makes for the best outcome, e.g., supposing well-being to be the sole relevant type of component, we can ask: Is the best outcome one in which overall well-being is maximized? Or may one outcome be better than another, even though it contains less well-being, so long as well-being is more equally distributed among subjects, or when the least well off are prioritized, or when all have enough?
4. How do we derive our evaluation of the subject from its relation to good outcomes, e.g., if the subject being assessed is acts, do we evaluate acts directly by reference to the extent to which they produce good outcomes? Or in some indirect way, such as their being compliant with a code of rules whose general acceptance would produce good outcomes?

In this paper, my central focus is on the first issue: What subject matter should consequentialists aim to address? Regarding the second and third issues, I intend what I say to be neutral across a wide variety of theories of what makes for good outcomes. Regarding the fourth issue, my focus will be on direct derivations of evaluations from good outcomes. However, indirect approaches such as rule consequentialism will figure briefly in our discussion.

1. CONSEQUENTIALISM AND RIGHT ACTION

The most well-known version of consequentialism is act utilitarianism. Its subject of evaluation is *acts*: it claims that the right act is the one that produces the best consequences. Its account of what makes outcomes good is *welfarist*: it claims that the best outcomes are those in which well-being or happiness is promoted. Its preferred distributive principle is *maximization*: the best outcome is the one in which well-being is maximized, no matter how that well-being is distributed across subjects. Finally, its evaluation of acts is *direct*: it assesses acts directly in terms of the goodness of their consequences (relative to the goodness of the consequences of the available alternative acts).

Let us begin then by considering consequentialist theories that, like act utilitarianism, are theories of right action, or of what we ought to do.¹ Such theories have significant initial plausibility, given that they embody the appealing thought that we should try to bring about what is good and try to eliminate what is bad. However, it is important to notice that talk of right action, or of what we ought to do, may be ambiguous. Consider the following two construals of the act consequentialist claim:

1. We are *morally obliged* to do whatever brings about the best (expected) consequences.² All other available actions are *morally wrong*.
2. What there is *most reason* for us to do is whatever brings about the best (expected) consequences. All reasons for action are grounded in production of the good.³

Formulation 1 faces very strong demandingness objections. It implies that, given the state of the world, where there is so much preventable suffering, we are morally obliged to devote almost all of our spare time, money, and energies to help-

- 1 A *fully fledged* act consequentialism perhaps is further committed to the view that what *makes* actions right is that they produce the best consequences. This stands in contrast, for example, to a divine-command theory that says that what makes actions right is that God commands them, but that God always commands those actions that produce the best consequences. See Heydt (“Utilitarianism before Bentham”) for the historical importance of such versions of consequentialism.
- 2 Whether we characterize a consequentialist claim in terms of the best consequences or the best expected consequences may be seen as a fifth relevant issue. For the purposes of this paper, I lay this question aside.
- 3 It is quite open to a theorist to assert the first claim without the second—to say that reasons for action can have other sources besides promotion of the good, but that the fact that some action will best promote the good is always a trumping factor. I focus here on the stronger view, which makes both claims.

ing strangers in need.⁴ If we do anything less than the most we can to help, we are acting morally wrongly. There are several aspects to such demandingness objections, each of which leaves this extremist form of act consequentialism looking seemingly untenable. First, it leaves us with very few morally permissible options; at every turn, we may only pursue one of those courses of action that will expectably bring about the most good. The options available will be generally unattractive, requiring me to forgo many of my most treasured projects, pastimes, and even relationships. Most of the time, a much better bet in maximally promoting the impartial good will rather involve working for effective charities, campaigning for trade justice, and so on. A second aspect of the extremist view's apparent implausibility is its condemnation of apparently morally admirable behavior as morally wrong. Someone who devotes a very substantial amount of her spare time and money to good causes, but who falls short of doing the most she could do is, by the lights of this view, failing in her moral obligations. A related unintuitive implication is that the view seems to leave no room for supererogatory action—action that is morally good, but that goes beyond what duty or obligation requires. Finally, the view suggests that we are only allowed to be partial toward our loved ones to a very limited degree—just to that degree that is mandated by an impartial calculus.

We should note that formulation 2 does not face the same objections. Standard demandingness objections apply specifically to accounts of moral obligation, not to mere rankings of the choiceworthiness of actions. To say that someone acts morally wrongly, that she has failed in her moral obligations, is not simply to say that she has done something other than the very best thing she could have done by the lights of morality, but to add a positive *criticism* of her acting as she does.

One common way of construing what is distinctive about judgments of moral obligation and moral wrongness is in terms of the distinctive sanctions of morality—those of the paradigmatic moral sentiments of blame and guilt. It is this feature that gives demanding theories of moral obligation their capacity to disturb—such theories imply that we must give up many of our treasured pastimes, projects, and relationships for the sake of aiding strangers, or else be deserving of distinctively painful feelings of blame and guilt.⁵

- 4 This conclusion is contingent, of course, on what account of the good is proposed. On any impartial account of the good that has a commonsensical conception of welfare, where intense suffering is counted as a significant bad, the conclusion seems unavoidable.
- 5 See Gibbard (*Wise Choices, Apt Feelings*, esp. ch. 3) and Skorupski (*The Domain of Reasons*, esp. ch. 12) for development and defense of the view that moral wrongness is centered around the sentiments of blame and guilt. Just how tight the connection is between moral wrongness and blameworthiness is a matter of dispute. In assuming in this paper that there

Formulation 2 has no such implications. It says simply that the action that has the best consequences is the one that has most to be said in its favor, without implying that if one does anything else, one merits criticism or moral sanction.⁶ It might be pressed that an analogous demandingness objection applies to formulation 2, namely that it implies that we must live a very self-sacrificial lifestyle on pain of meriting charges of *irrationality*. But this charge is under-motivated. Formulation 2 does indeed imply, given a reasonable account of the good, that unless we do the most we can to help those in need, we are falling short of doing the very best we can by the lights of reason. But this is a charge we can happily live with. The life of the moral saint may be one that has the most to be said for it overall, the one that gets the strongest endorsement from the standpoint of reason. But one does not merit positive charges of irrationality, in the ordinary, rhetorically loaded sense of that term, simply for acting otherwise.⁷ One's behavior may be perfectly *understandable* (and thus not *irrational*) if grounded in good reasons, even when those reasons are less strong than reasons to do something that will be difficult or costly to one's own well-being.

Our first conclusion then is that if consequentialism is to be understood as

is indeed a significant connection between the two, I am *not* assuming the very controversial view that there can never be a gap between the wrong and the blameworthy. For instance, we might confidently judge some action (say, insulting a colleague) to be morally wrong, but think the agent is not blameworthy because of some extenuating circumstance (she had just received extremely upsetting news). In another type of case, we might judge that some action is "objectively" wrong, but the agent is not blameworthy because she is in no epistemic position to know the features of the action that make it a bad idea (for example, she is in no position to know that putting the plug in this socket will result in someone getting a nasty electric shock). In both these sorts of cases, even though moral wrongness and actual blameworthiness come apart, it still seems natural to think there is a close connection between the judgment that the act is wrong and *hypothetical* blameworthiness. In the first case, part of what is meant by saying the action is wrong is that it is the sort of thing that one would merit feelings of blame for doing *absent* extenuating circumstances. In the second case, in saying that the action is "objectively" wrong, we are committing to the idea that one would merit feelings of blame if one was fully informed of the relevant facts and yet performed that action.

6 It is perhaps more common in everyday talk to speak simply of what we should do, rather than of what there is most reason to do. I deploy the latter formulation simply to avoid the implications of meriting blame or serious criticism that can come with "should" and more frequently with "ought" in certain contexts. It is noteworthy that when we evaluate possible courses of action, we very frequently use very close synonyms for "reason" and "most reason," e.g., "There's something to be said for staying a bit longer, but I think there's most to be said for leaving now." Furthermore, in the case of ordinary normative talk about beliefs and feelings, which I discuss below, we tend to use the language of "reason" and "most reason" very frequently.

7 For further discussion, see McElwee, "Impartial Reasons, Moral Demands."

a direct theory of right action, or of what we ought to do, then this is best construed as a claim about what we have most reason to do, not as a theory of our moral obligations.⁸

2. CONSEQUENTIALISM AS A GENERAL THEORY OF NORMATIVITY

Consequentialism may be best known as a theory assessing actions. But why focus solely on actions? Other things besides actions require ethical evaluation. And other things besides actions can have good or bad consequences. Consequentialists have thus been led to make distinctively consequentialist assessments of many other things besides acts—in particular, motives, character traits, moral rules, moral codes, and sets of institutions. As Jeremy Bentham says, “It is with disposition [of character] as with everything else: it will be good or bad according to its effects: according to the effects it has in augmenting or diminishing the happiness of the community.”⁹

The idea that everything should be assessed in consequentialist terms has come to be known as global consequentialism, a term coined by Philip Pettit and Michael Smith:

Global consequentialism identifies the right x , for any x in the category of evaluands—be the evaluands acts, motives, rules, or whatever—as the best x , where the best x , in turn is that which maximises value.¹⁰

Similarly, Shelly Kagan writes that the “most plausible version of consequentialism will be direct with regard to everything.”¹¹

Derek Parfit makes explicit that consequentialism is to cover all possible evaluands:

Consequentialism covers, not just acts and outcomes, but also desires, dispositions, beliefs, emotions, the colour of our eyes, the climate, and everything else. More exactly, C covers anything that could make outcomes

8 A useful comparison here is with the “scalar” consequentialism defended in Norcross, “The Scalar Approach to Utilitarianism,” and Crisp, *Reasons and the Good*. Such a view endorses a consequentialist ranking of actions from best to worst (or from what there is most reason to do to what there is least reason to do). But what is distinctive about the scalar view is not its consequentialist ranking, but its rejection of any supplementary account of moral obligation. I argue against such a rejection in McElwee, “Consequentialism and Permissibility” and “Should We De-Moralize Ethical Theory?”

9 Bentham, *An Introduction to the Principles of Morals and Legislation*, 246.

10 Pettit and Smith, “Global Consequentialism,” 121.

11 Kagan, “Evaluative Focal Points,” 151.

better or worse. According to C, the best possible climate is the one that would make outcomes best.¹²

What attractions might there be in adopting such a global consequentialism? Julia Driver notes that one prominent rationale for the view comes in response to the virtue ethical challenge to consequentialism.¹³ Consequentialists, along with other modern moral theorists, are accused of giving exclusive attention to the evaluation of action, particularly couched in terms of the morally obligatory, at the expense of attention to motives and character. One consequentialist response to this is to assert that consequentialism has a ready-made way of evaluating motives and character traits; this can be done in just the same way as actions are to be evaluated, directly in terms of the consequences they (tend to) produce. Another attraction Driver claims for global consequentialism is that it gives an attractive account of what she calls “normative ambivalence,” where “a stable evaluation, or a unitary evaluation, is hard to achieve because we are really thinking about two different things: [for example] the agent’s action and the character the agent is expressing through the action.”¹⁴ As Robert Adams notes, the consequentially optimal set of motives for an agent to have may lead that agent on occasion to act in a consequentially suboptimal way.¹⁵ In such cases, we may feel a tension in our normative assessments, at the same time wanting to endorse the agent’s behaving as she does and yet wanting to criticize it. But this tension, Driver suggests, can be relieved simply by making a twofold claim, offering a direct consequentialist account of the action (it is wrong because it has overall bad or suboptimal consequences) and another direct consequentialist account of the set of motives in play (it is right because it has better overall consequences than any other set of motives, even if in this case it leads to wrong actions). Global consequentialism thus allows us to say all we want to say.

The focus of Pettit and Smith’s discussion is to argue that this global consequentialism is preferable to “local” forms of consequentialism. Local forms of consequentialism give a direct consequentialist account of a privileged evaluand, and then give an assessment of other evaluands by reference to some relation they stand in to the privileged one. Rule consequentialism is an example of this local consequentialist pattern: its privileged evaluand is sets of moral rules. The

12 Parfit, *Reasons and Persons*, 25.

13 Driver, “Global Utilitarianism.”

14 Driver, “Global Utilitarianism,” 169.

15 Adams, “Motive Utilitarianism.”

right set of moral rules is that whose general acceptance produces the most good. A right action is one that complies with this right set of rules.¹⁶

Pettit and Smith argue that global consequentialism is to be preferred to *all* forms of local consequentialism. After all, why should we privilege any particular evaluand? Local consequentialisms will unduly ignore those consequences of each evaluand that are not mediated by the privileged evaluand. We should, for example, reject a local consequentialism that defines the right act as one that has the best consequences, but then defines the right motives or rules as those that lead to the promotion of right acts, on the following basis:

Someone's possession of certain motives, or his or her having internalised certain rules, may have consequences that are not mediated by any act to which those motives or rules give rise. Your clear benevolence towards me, and mine towards you, can provide each of us with a sense of warmth and reassurance independently of any acts that it occasions. And the mere knowledge that you have internalised a rule of promise-keeping provides me, well in advance of any contract we enter into, with a rich sense of the arrangements we may form.¹⁷

However, in order to establish that global consequentialism is the most plausible version of consequentialism, it is not enough to establish that it is more plausible than local consequentialisms that give a direct consequentialist treatment of a privileged evaluand and then supplement that with accounts of other evaluands that are derived from the privileged one. One would also need to show that global consequentialism is to be preferred to theories that are directly consequentialist about actions (or some other particular evaluand), and then complement that claim with a *non*-consequentialist account of other evaluands—i.e., an account that is neither directly consequentialist, nor one that assesses these other evaluands by reference to some relation they stand in to the privileged evaluand.

In fact, global consequentialism comes off badly when compared to some such “hybrid” views. I will put the point first in terms of reasons, before going on to address global consequentialists on their own terms, in section 4. I have suggested that consequentialism has some attraction as a theory of what we ought to do, specifically when interpreted as the claim that what there is *most reason* to do is what brings about the best consequences. We can apply the global consequentialist's challenge—“Why privilege actions?”—to this formulation. If

16 For detailed discussion of different possible formulations of rule consequentialism, see Hooker, *Ideal Code, Real World*.

17 Pettit and Smith, “Global Consequentialism,” 122.

consequentialism is a plausible theory of reasons for action, might it not equally be a plausible theory of reasons for other reason-responsive states—of reasons for belief, and of reasons to feel?

Let us take first the claim as applied to beliefs:

There is most reason to believe whatever belief will bring about the best consequences. There is reason to believe *B* just to the extent that believing *B* will bring about some good.

This consequentialist account of reasons to believe looks very unpromising.¹⁸ Consider a pair of detectives investigating a murder. In working out whether there is reason to believe Jones committed the crime, they should (and detectives typically do) reflect on the evidence available to them. Is there evidence that Jones was at the scene of the crime at the time the crime was committed? Or evidence that he was somewhere else? Is there evidence that Jones had some motive to commit the crime? Or evidence that he had no such motive?

In working out whether there is reason to believe that Jones committed the crime, they should not (and detectives typically do not) start thinking about the goodness of the consequences of their having various beliefs about whether Jones committed the crime. Suppose one detective says, “We’ve already got Jones locked up in a cell. If I were to believe that Jones committed the crime, I’d be able to sleep better tonight, believing that the killer is behind bars. So there is good reason to believe Jones committed the murder.” This would seem patently absurd—this is not the way our ordinary talk of reasons to believe operates.¹⁹ Rather, reasons to believe are based on evidence. A direct consequentialist account of reasons to believe, evaluating beliefs in light of the same goods used to evaluate actions, seems implausible.²⁰

Might we not be more tempted in some other cases to say that the conse-

18 At least if it simply applies the same account of the good being used to assess reasons to act. Perhaps a more plausible account of reasons to believe that may yet be described as consequentialist is one in which there is a *sui generis* good in light of which we should evaluate beliefs, distinct from the good in light of which we should evaluate actions (e.g. well-being). For example, it might be suggested that the belief there is most reason to believe is the one that maximizes overall true belief, or significant true beliefs, or the ratio of true beliefs to false ones. This sort of view would be quite different from a unified consequentialism, with a single account of the good, of the kind that global consequentialist writers clearly have in mind, and I do not aim to make any assessment of it in this paper. See Dunn, “Epistemic Consequentialism.”

19 Note that the absurdity does not lie in this being a suboptimal “decision procedure” for establishing what there is reason to believe. Rather, it lies in a misidentification of what the reason-making factors are.

20 Chappell (“Fittingness”) similarly observes that the case of beliefs (and that of feelings)

quences of believing should bear on what we ought to believe? An opponent of the view under discussion might respond by saying, “Of course the consequences of having beliefs matters. Believing (unwarrantedly) that you are the best man for the job can help you get the job. So you ought to try to cultivate such a belief.”

Insofar as we think reasons for *action* are provided by the consequences of our actions, that final claim about what we ought to *do* may be absolutely correct. Our beliefs are not generally under the immediate control of our will, but there are courses of action we can take in order to cultivate certain sorts of beliefs and to suppress others. When it is possible for us, at sufficiently little cost, to cultivate beliefs that will produce good consequences, we plausibly have reason to do so; there are frequently reasons to *bring it about* that I have certain beliefs. But these reasons are reasons for *action*.

The natural thing to say about such cases, I think, is that there is no reason to believe that I am the best man for the job (if all my evidence suggests otherwise) but that there may be practical reason to try to bring it about that I believe I am, if doing so is really going to bring about some good consequence. Some may prefer to talk instead about *two types of reasons to believe*: reasons of evidential warrant and consequence-based reasons. Perhaps not that much hangs on whether we call the latter as well as the former reasons for belief. What is more important is that there seems to be no intelligent composite question of what I have most reason to believe overall, which somehow combines reasons of the two kinds. Instead, there is the practical question of what beliefs I ought to try to cultivate, and there is the distinct question of what beliefs are epistemically warranted.

In any case, what is relevant for present purposes is the relatively weak claim that *not all* reasons for belief are grounded in the consequences of those beliefs. A *purely* consequentialist account of reasons to believe seems clearly implausible.

Why is this important? Whoever thought that consequentialism was supposed to provide an account of reasons to believe? Well, first we seem to have established that some evaluands lend themselves more readily to a consequentialist treatment than others—*contra* the global consequentialist’s claim. Notwithstanding the fact that beliefs can have good and bad consequences, such features clearly do not exhaust the considerations bearing on their normative status. The import becomes much more significant, however, when we turn to the third category of reason-responsive states—namely, feelings.

Consider the claims:

looks especially jarring for global consequentialists in their attempt to extend consequentialism beyond its proper domain of action.

There is most reason to feel whatever feeling will bring about the best consequences. There is reason to feel *F* just to the extent that feeling *F* will bring about some good.

A consequentialist account of reasons to feel looks just as unattractive as a direct consequentialist account of reasons to believe. Natural, appealing claims about reasons to feel, making no reference to the consequences of having the feeling, abound:

There is reason to feel sad when you are bereaved.

There is reason to feel grateful when someone does you a good turn.

There is reason to feel disappointed when you lose a cup semifinal.

There is reason to feel relief when you are cured of a dangerous illness.

We do not need to know that feeling a certain way will have good consequences in order to know what there is reason to feel. Reasons to feel appear to be determined by the fittingness of the feeling to the object of the feeling, rather than by the good or bad consequences of having the feeling.²¹

Again, an opponent might respond at this point by saying, “Of course the consequences of having feelings matters. Some feelings are very harmful. Feelings of envy, for instance, can eat away at you and cause misery to you and your loved ones. We ought to try to eliminate such feelings.” And again, if we think reasons for *action* are provided by the consequences of our actions, the claim about what we ought to *do* may be quite correct. As in the case of belief, there are courses of action we can take in order to cultivate certain types of emotional responses and to suppress others. When we are able to costlessly cultivate feelings that will produce good consequences, we have some reason to do so. There are *practical* reasons to bring it about that I feel some way or another.

This sort of distinction has figured in discussions of fitting attitude analyses of value, under the heading of the “wrong kind of reasons.” Just because it will have good consequences if I feel admiration for some cruel and powerful tyrant, this does not mean that the tyrant is admirable. The most natural-sounding thing to say, I think, is that there is no reason to admire the tyrant (he is not admirable) but that there is practical reason to try to bring it about that I admire him, if doing so is, for example, going to spare me a painful death at the hands of

21 Again, it might be suggested, in a “consequentializing” spirit, that there is a *sui generis* good in light of which we should evaluate feelings, distinct from the good appropriate for assessing acts. For example, we have most reason to feel whatever feeling best maximizes overall fittingness between one’s feelings and their objects. I do not intend to evaluate this sort of view in this paper.

his henchmen.²² Instead of talking about *two types of reasons to feel* (reasons of fittingness and consequence-based reasons), it seems more natural to talk about (fitting) reasons to feel, and practical reasons to cultivate feelings. But again, it is more important to recognize that there is no composite question of what I have most reason to feel overall, combining reasons of the two kinds. Instead, there is simply the practical question of what feelings I ought to try to cultivate, and a distinct question of what feelings are fitting, or merited, or apt, or appropriate.

As before, what is most relevant here is the weaker claim that *not all* reasons for feelings are grounded in the consequences of those feelings. A *purely* consequentialist account of reasons to feel seems clearly implausible. Our second conclusion, then, is that consequentialism is *not* a plausible theory of reasons in general.

3. REASONS TO FEEL AND ETHICAL EVALUATION

The correct treatment of reasons to feel has wider importance than is generally appreciated. Many of our ethical (and more generally evaluative) questions are, in significant part, matters of what there is reason to feel.

3.1. *Moral Obligation*

Take the case of moral obligation, already addressed briefly in section 1. Consequentialists frequently claim that we are morally obliged to do whatever brings about the best (expected) consequences. We noted that such claims face strong demandingness objections, on any substantive account of the morally obligatory, which understands this as implying more than that an action is morally best. Why does it seem so objectionable to say that we are *morally obliged* to devote our lives to helping distant strangers, that we would be acting *morally wrongly* if we did otherwise? Its objectionableness is plausibly explained by the sentimental core of charges of moral wrongness and of violating moral obligations. When a moral theory claims merely that a life of extreme altruism would be *morally best*, we do not typically reply, “That’s too demanding to be morally best.” But when a moral theory says that the life of extreme altruism is *morally required* or *obligatory*, we confidently judge that it is too demanding. What is added by claims about what is morally obligatory is the imputation of blame or guilt for failure to act in the recommended way. Again, this is why the extreme act consequentialist claim is so unsettling—it says we must give up our relatively comfortable lifestyles or else be such as to merit the sanctions of the moral sentiments.

Consider once more the agent who does a very substantial amount to

22 Rabinowicz and Rønnow-Rasmussen, “The Strike of the Demon.”

help needy strangers—devoting perhaps a third of his spare income and three nights a week of his spare time to helping the poor. It may be true of him that he could be doing even better, but it seems absurd to accuse him of acting in a way that is morally wrong. The reason why it sticks in the throat to say that he is acting morally wrongly, that he has failed in his moral obligations, is that he seems clearly not to merit feelings of blame for the level of altruism embodied in his behavior. His altruistic efforts in fact seem to merit admiration. If this line of thinking is correct, then our best account of our moral obligations will be shaped by our best judgments about when there is reason to feel blame toward an agent. It will not be a maximizing account like the extremist version of act consequentialism.

An alternative consequentialist account of moral obligation might endorse the connection between moral obligation and blameworthiness, but then go on to offer a distinctively consequentialist account of norms for blaming. One way to do this would be to give a direct consequentialist account of when we have most reason to perform distinctive blame *actions*, such as criticizing, remonstrating, shunning, or formally punishing. (Such an account will be of a piece with a more general direct consequentialist account of reasons for action.) But this is very jarring as an account of blameworthiness. Even if we accept the (controversial) view that we should perform such blame actions just when they will produce the best consequences, our judgments of blameworthiness seem tied not to the expediency of such blame actions, but instead to when there is reason to *feel* blame *feelings* toward the agent.²³ And, as argued in section 2, reasons to feel are more plausibly treated according to fittingness considerations, rather than in terms of the consequences of having the feelings in question.²⁴

3.2. *Virtue*

A second example of a central ethical concept that must plausibly take account of reasons to feel is virtue. Julia Driver, following Bentham and Mill, argues that virtues are character traits that are systematically instrumental in promoting

23 And so there may be (consequence-based) reason to shun or punish someone who is not blameworthy, and there may be (consequence-based) reason to forgo shunning or punishing someone who is blameworthy.

24 It is beyond the scope of this paper to give an account of when it is fitting to feel blame toward an agent for acting as she did. But such an account will generally need to give significant attention to the value of the consequences of the act being assessed. All else being equal, one is frequently blameworthy for performing actions with bad consequences, and not generally blameworthy for performing actions with good consequences.

good consequences.²⁵ But this view fits awkwardly with our intuitive judgments about virtue.

Virtues do generally produce good consequences in the actual world, so the clearest-cut examples where this instrumentalist treatment seems wrong may be rather unusual. But any correct account of what makes for virtue must cover unusual counterfactual cases. Suppose a powerful demon made it the case that malice systematically produced good consequences. Would this really suffice to make malice—*wishing suffering upon others*—virtuous? This rings false. A plausible explanation why is that our judgments of virtue are in large part governed by judgments about reasons to feel—in this case, about what traits there is reason to admire. Those traits that we acclaim as virtues are not necessarily those that have the best consequences, but those that we have reason to morally admire. We should not regard malice as a virtue even in these strange circumstances, because malice is not admirable—there is no good reason to feel admiration for malicious people. Rather, the circumstances are merely ones where, unusually, we have strong *practical* reason to cultivate the vice of malice.

More homely examples make a similar point, though perhaps less starkly. Consider the disposition to obey the law from fear of punishment. If this trait is likely to be more reliable in getting people to obey (good) laws, then it is a disposition we have good practical reason to cultivate, but it rings hollow to call it a virtue because it is not an especially admirable trait to possess. There may be more noble motivations to obey the law, which we do have reason to admire and that we would thus more readily describe as virtuous. But if the former are significantly more reliable than the latter in promoting the good, then we have good reason to cultivate the former in ourselves and others, given the importance of the goods at stake.

On the present proposal, there being good practical reason to cultivate a trait in prevailing circumstances is not sufficient to make the trait a virtue; rather, we only acclaim a trait as a virtue if there is reason to feel admiration toward the agent who has the trait. This account still leaves open what substantive conception of virtue to adopt. But a strong candidate view that fits well with the proposal is the Aristotelian one that virtuous traits are ones that involve correct responsiveness to value—traits that involve “loving the good,” as Thomas Hurka puts it, or “being for the good,” as Robert Adams calls it.²⁶ We morally admire, and thus call virtuous, those traits that embody such correct responsiveness to value, rather than those that simply lead to good consequences. And such cor-

25 Driver, *Uneasy Virtue* and “The Virtues and Human Nature.” See Skorupski (“Externalism and Self-Governance”) for critical discussion along the lines pursued here.

26 Hurka, *Virtue, Vice and Value*; Adams, *A Theory of Virtue*.

rect responsiveness to value itself partly consists in feeling what there is (fitting) reason to feel: taking pleasure in, for example, the happiness of others and being pained by their suffering.

This very natural view about what underlies common moral assessments goes some way toward undercutting Driver's suggestion that global consequentialism best explains the phenomenon of "normative ambivalence." The sort of case Driver has in mind is one where some character trait, or set of motives, that we would generally endorse leads an agent to perform an action other than the one that we would wish them to perform. So, for example, the compassionate person fails to take the opportunity to kill an evil dictator because she cannot bring herself to ruthlessly strangle him with a shoelace.²⁷ Driver is surely correct to say that such cases are best treated by giving separate evaluation of the act and of the operative motives or character traits. But even if we endorse a direct consequentialist evaluation of the act, this does not mean that a direct consequentialist evaluation of the motive or trait is most plausible. A deep-seated dislike of violently inflicting pain and suffering is *fitting*, not just instrumentally good. Talk of doing the right thing for the wrong reason or from a bad motive, and of doing the wrong thing for the right reason and from a good motive is a staple of common moral judgment. But talk of good motives is most naturally construed in terms of motives that aim at the good, not in terms of their being instrumental in producing the good.

That some character traits involve feelings and actions that constitute intrinsically fitting responses to value is something that consequentialists really ought to accept but often overlook in offering a purely instrumentalist treatment of virtue. Take the utilitarian axiology that happiness is good and suffering bad. Surely part of what is involved in making this claim is commitment to the idea that it is fitting to desire, to approve, to take pleasure in the production of happiness, and likewise fitting to abhor, to disapprove of, to be pained by suffering. We need not await some further judgment about the consequences of having such attitudes to know that there is something correct about them, that they enjoy some positive normative status. Such responses can themselves be evaluated instrumentally—they can be good as means to securing happiness. But this does not exhaust their normative import. Consequentialists have generally eschewed talk of fitting attitudes, perhaps because it seems to open the door to other sorts of values—"intuitionist" values of the kind that Mill was combating in the nineteenth century. But a consequentialism about reasons for action can fit happily with talk of fitting feelings.²⁸

27 Driver, "Global Utilitarianism," 172–73.

28 See Hurka (*Virtue, Vice and Value*) for a form of consequentialism that makes room for ap-

Our third conclusion is that offering plausible accounts of reasons to feel is extremely important because many ethical judgments are partly constituted by judgments about what there is reason to feel.

4. GLOBAL CONSEQUENTIALISM

Let us return now to global consequentialism. As we saw earlier, Pettit and Smith characterize the view in terms of *rightness*:

Global consequentialism identifies the right x , for any x in the category of evaluands—be the evaluands acts, motives, rules, or whatever—as the best x , where the best x , in turn is that which maximises value.²⁹

So, in summary, the right x is the x that maximizes value.

It is useful to look at how talk of the “right x ” figures in ordinary discussion for different instances of x . Consider again those cases where x is a reason-responsive state, but not an act:

What’s the right thing to believe about this?

What’s the right way to feel about this?

When phrases like these crop up in everyday talk, they pretty clearly refer to what there is reason (fitting reason) to believe or to feel. It would ordinarily be decidedly odd to start talking about the consequences of having the belief and of having the feeling in response to such questions couched in terms of rightness. The right thing to *do* may in unusual circumstances be to bring about an unfitting belief (e.g., where the evidentially warranted belief is very distressing) or an unfitting feeling (e.g., where having the fitting feeling will lead to my being tortured). But the most natural construal of the question of the “right belief” and the “right feeling” is in terms of, respectively, reasons to believe (in the sense of what is fitting to believe) and reasons to feel (in the sense of what is fitting to feel).

Global consequentialists most often discuss motives, character traits, and codes of moral rules as the x to be given a consequentialist treatment. But we have already seen that motives and character traits plausibly involve feelings that may be fitting or unfitting, so talking of the “right motive” and the “right charac-

appropriate response and characterizes virtue in these terms. See Chappell (“Fittingness”) for excellent discussion of how consequentialists can benefit by making fitting attitudes central to their theory.

29 Pettit and Smith, “Global Consequentialism,” 121.

ter trait” as if these statuses were settled in a purely consequentialist way is again likely to be jarring.

4.1. *Motives*

Consider a scenario where the right thing for Bob to do is to help Andrea. We might ask what is the right motive for Bob to have and to act upon. Compare these two motives:

1. Bob helps Andrea out of compassion.
2. Bob helps Andrea out of a desire to humiliate her lover.

There are possible circumstances in which it would have the best outcome for Bob to act out of the latter motive—perhaps this will have a much higher probability of Bob *successfully* helping Andrea if he acts from this motive, and for once, the malicious motive is unlikely to have significant further effects. But does this really settle the question of whether it is the *right* motive? A relevant normative feature of the two motives seems to have been ignored. On the view outlined above, motives can be intrinsically fitting, involving correct responsiveness to value. In the present example, the first motive involves being pained by something bad, Andrea’s suffering, and desiring something good, the relief of her suffering. The second motive involves taking pleasure in and desiring something bad, the humiliation of Andrea’s lover. Given this, I think we would more naturally describe the first motive as the “right” one, notwithstanding its inferior consequences. “Rightness” talk, like reasons talk, seems to go more readily with fittingness considerations than with consequence-based ones. At the very least, fittingness considerations seem a second relevant normative feature of motives alongside the consequences of having the motive.

4.2. *Character Traits*

As we saw in the discussion of virtue, above, it seems that similar things should be said about character traits. Talk of the “right character traits” could be construed as referring to the character traits we have reason to cultivate (because, on the consequentialist view, their cultivation will best promote the good). But it is at least as naturally interpreted as referring to those traits that involve fitting motives, fitting feelings, fitting responses to value. We might thus be tempted to say that the “right” character trait for Bob to have is compassion, rather than a determined one-upmanship, even when the latter is just as effective, or even more effective, in leading him to act such as to promote the good. The instrumentalist about the evaluation of character traits may respond that compassion is to be preferred to these latter traits because it is in general more likely to lead to the

promotion of the good. But again this seems to leave out something extremely important from our assessment.

4.3. *Codes of Moral Rules*

Similar observations apply to codes of moral rules, when we ask which is the “right” code of moral rules. We should begin by asking what a code of moral rules is. It is, at least in part, a collection of judgments of the form, “One is morally obliged to do x in circumstances C_1 ,” “One is morally obliged to do y in circumstances C_2 ,” and so on. But our best judgments about what is morally obligatory, I have suggested, involve judgments about when there is reason to *feel* blame toward agents who fail to comply. So we might naturally construe the phrase “the right code of moral rules” as shorthand for talking about the right way to feel about someone who fails to do x in circumstances C_1 , the right way to feel about someone who fails to do y in circumstances C_2 , and so on. And as noted above, these seem to be judgments that most convincingly admit of non-consequentialist treatments.

If we ask the *practical* question about which code of moral rules we should we try to inculcate in ourselves and others, then perhaps in some circumstances we have good reason, grounded in the good consequences of so doing, to inculcate a code that involves treating as morally obligatory something that is not morally obligatory; we should inculcate patterns of feeling that include occasional feelings of blame toward those who do not merit such feelings, and perhaps even *beliefs* that there is reason to feel blame toward someone whom there is no reason to feel blame toward. But saying that this code is the “right” one, just in virtue of its inculcation having the best consequences in certain circumstances, is at best misleading. What is essential is to separate out the practical question, which admits of a plausible consequentialist answer, and the various questions about what there is reason to feel, which do not admit of a plausible consequentialist answer.

So in the global consequentialist’s favorite cases—motives, character traits, moral codes—a direct consequentialist treatment of the “right x ” seems most plausible insofar as the specific evaluative question being addressed is reducible to a question about what there is reason to do. But even in those cases, the most natural construal of the question about the “right x ” is not the practical one, but one that involves questions about reasons to feel, about fitting feelings, which do not admit of a plausible consequentialist treatment.³⁰

30 What about those evaluands that seem to involve no reason-responsive state whatsoever? Driver notes the sheer oddity of talking of the right eye color or the right climate: “This is odd because these objects are not agents, and we tend to intuitively restrict moral evalua-

Our fourth conclusion then is that the global consequentialist claim is at best misleading. It claims that, for x in general, the right x is the one that results in the best outcomes. But this is most plausible when the “right x ” is construed as the “ x there is reason to act to bring about.”³¹ So again, consequentialism appears most plausible simply as a theory of reasons for action.³²

5. THE PRESSURE FROM SENTIMENTALISM

My aim in this paper has not been to offer positive support for consequentialism. Rather, I have tried to establish that a form of consequentialism that limits itself to being a theory of reasons for action, and that complements this with non-consequentialist accounts of other normative questions, is more plausible than one that gives a direct consequentialist account of everything.³³ That is not to say that any form of consequentialism is ultimately convincing. Indeed, one might think that a very plausible competitor view emerges naturally from our discussion—one in which practical reasons have two sources, the good and the sentiments.

I suggested above that in order to know what reasons to feel there are, we need not know the consequences of having those feelings. It is sufficient for knowing that there is (at least some) reason to feel gratitude toward Tom that I know that Tom has done me a good turn. I do not need to investigate whether feeling grateful to Tom will have some good effect. Might we not then be tempt-

tion to features relevant to agency. . . . Moral agents are sensitive to reasons; climates are not” (“Global Utilitarianism,” 173). One way of making sense of such judgments of the “right climate” or the “right eye color” is to read them as claims about which is the climate or eye color that it is right to choose when there is reason to choose between climates or between eye colors. Again, this interpretation of the global consequentialist’s claims simply makes them instances of direct act consequentialism, claims specifically about reasons for *action*. For further discussion of such cases, and their relation to the “ought implies can” principle, see Streumer, “Can Consequentialism Cover Everything?”; Brown, “Blameless Wrongdoing and Agglomeration”; and Streumer, “Semi-Global Consequentialism and Blameless Wrongdoing.”

- 31 Perhaps, more precisely, the claim is most plausible when the “right x ” is construed as the “ x there is most reason to bring about *when possible*,” as there may be cases where the “right x ” seems one that we have (fitting) reason to will, wish for, or desire—it is the “optimal x ”—but would be impossible to bring about.
- 32 The arguments offered here buttress those offered in Chappell (“Fittingness”) for a similar conclusion.
- 33 I have offered no explicit argument that such a view is to be preferred to rule consequentialism. Though see Pettit and Smith (“Global Consequentialism”) and Kagan (“Evaluative Focal Points”) for arguments against forms of rule consequentialism that have an ultimately consequentialist justification.

ed to say just the same thing about *actions* that are expressions of reason-supported feelings? Why not say that the fact that Tom has done me a good turn is sufficient to establish that I have (some) reason to *thank* Tom? I need not inquire into whether thanking Tom will have some good consequence to know that there is at least some reason to thank him, even if that reason were to be outweighed by some bad consequences of thanking him. Intuitively, some actions—those that constitute expressions of feelings—can seem fitting in just the same way that feelings can.

We may conclude that just as there is reason to feel blame or resentment toward someone if they cheat you, so there is reason to *remonstrate* with them or to *protest*. Just as there is reason to admire your performance if it shows great skill, so there is reason to *applaud* you.

This idea has been articulated by John Skorupski, who defends what he calls the *bridge principle*.³⁴

Whatever facts give x reason to feel F give x reason to do the F -prompted action, in virtue of being a reason to feel F .³⁵

There is significant appeal to saying that some reasons to perform actions that constitute expressions of feelings are grounded in the fact that there is reason to feel those feelings, while other reasons for action may be based in the good consequences of the action.

One option for the consequentialist is to concede that there are indeed reasons for action grounded in the sentiments, and not in good consequences, but to insist that nonetheless there is always most reason to do that that brings about the most good. On this view, if one knows that some action will bring about the most good, then one can safely conclude that that is what one should do.

The relative pros and cons of these views will need to be addressed in future work.³⁶ My aim here has simply been to establish what consequentialism is most

34 On Skorupski's view, reasons for action come from *three* distinct sources. Some reasons for action are consequence-based reasons, grounded in the good; others are grounded sentimentally, via the bridge principle; and some are grounded in rights, by the demand principle, which claims that an agent has reason to do that which some person may permissibly demand of him. See Skorupski, *The Domain of Reasons*, pt. III, for a full discussion, and "The Triplism of Practical Reason" for a summary of the view.

35 Skorupski, "The Triplism of Practical Reason," 129.

36 A second important way in which a non-consequentialist account of reasons to feel may put pressure on a purely consequentialist account of reasons for action concerns sentiments of blame. Reasons to feel blame interact directly with reasons to act: a judgment that there is reason to feel blame toward some agent A appears to presuppose that A had stronger reason to do other than he did.

plausibly a theory of. A consequentialism more limited in its ambitions, which makes room for non-consequentialist answers to some normative questions, looks much more promising than one that attempts to provide a direct consequentialist answer to every normative question.³⁷

University of Southampton
b.m.mcelwee@soton.ac.uk

REFERENCES

- Adams, Robert Merrihew. "Motive Utilitarianism." *Journal of Philosophy* 73, no. 14 (August 1976): 467–81.
- . *A Theory of Virtue: Excellence in Being for the Good*. Oxford: Oxford University Press, 2006.
- Bentham, Jeremy. *An Introduction to the Principles of Morals and Legislation*. Oxford: Blackwell, 1948.
- Brown, Campbell. "Blameless Wrongdoing and Agglomeration: A Response to Streumer." *Utilitas* 17, no. 2 (July 2005): 222–25.
- Chappell, Richard Yetter. "Fittingness: The Sole Normative Primitive." *Philosophical Quarterly* 62, no. 249 (October 2012): 684–704.
- Crisp, Roger. *Reasons and the Good*. Oxford: Oxford University Press, 2006.
- Driver, Julia. "Global Utilitarianism." In *The Cambridge Companion to Utilitarianism*, edited by Ben Eggleston and Dale E. Miller, 166–76. Cambridge: Cambridge University Press, 2014.
- . *Uneasy Virtue*. Cambridge: Cambridge University Press, 2001.
- . "The Virtues and Human Nature." In *How Should One Live? Essays on the Virtues*, edited by Roger Crisp, 111–30. Oxford: Oxford University Press, 1996.
- Dunn, Jeffrey. "Epistemic Consequentialism." *Internet Encyclopedia of Philosophy*. <https://www.iep.utm.edu/epis-con>.
- Gibbard, Allan. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Oxford: Oxford University Press, 1990.
- Heydt, Colin. "Utilitarianism before Bentham." In *The Cambridge Companion*

37 Early versions of this paper were presented at the University of St. Andrews, and at Senate House, University of London at a symposium in honor of John Skorupski. I am very grateful to the participants at each of these for their feedback. I would also like to thank David Plunkett, Ben Sachs, Justin Snedegar, and Richard Yetter Chappell for very helpful comments on drafts of the paper.

- to *Utilitarianism*, edited by Ben Eggleston and Dale E. Miller, 16–37. Cambridge: Cambridge University Press, 2014.
- Hooker, Brad. *Ideal Code, Real World*, Oxford: Oxford University Press, 2001.
- Hooker, Brad, Elinor Mason, and Dale E. Miller, eds. *Morality, Rules, and Consequences: A Critical Reader*. Edinburgh: Rowman and Littlefield, 2000.
- Hurka, Thomas. *Virtue, Vice and Value*. Oxford: Oxford University Press, 2001.
- Kagan, Shelly. “Evaluative Focal Points.” In Hooker, Mason, and Miller, *Morality, Rules, and Consequences*, 134–55.
- McElwee, Brian. “Consequentialism and Permissibility.” *Utilitas* 22, no. 2 (June 2010): 171–83.
- . “Impartial Reasons, Moral Demands.” *Ethical Theory and Moral Practice* 14, no. 4 (August 2011): 457–66.
- . “Should We De-Moralize Ethical Theory?” *Ratio* 23, no. 3 (September 2010): 308–21.
- Norcross, Alastair. “The Scalar Approach to Utilitarianism.” In *The Blackwell Guide to Mill’s Utilitarianism*, edited by Henry R. West, 217–32. Oxford: Wiley-Blackwell, 2006.
- Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press, 1984.
- Pettit, Philip, and Michael Smith. “Global Consequentialism.” In Hooker, Mason, and Miller, *Morality, Rules, and Consequences*, 121–33.
- Rabinowicz, Wlodek, and Toni Rønnow-Rasmussen. “The Strike of the Demon: On Fitting Pro-Attitudes and Value.” *Ethics* 114, no. 3 (April 2004): 391–423.
- Skorupski, John. *The Domain of Reasons*. Oxford: Oxford University Press, 2010.
- . “Externalism and Self-Governance.” *Utilitas* 16, no. 1 (March 2004): 12–21.
- . “The Triplism of Practical Reason.” *Ratio* 25, no. 2 (June 2012): 127–47.
- Streumer, Bart. “Can Consequentialism Cover Everything?” *Utilitas* 15, no. 2 (July 2003): 237–47.
- . “Semi-Global Consequentialism and Blameless Wrongdoing: Reply to Brown.” *Utilitas* 17, no. 2 (July 2005): 226–30.

ERROR THEORY, UNBELIEVABILITY, AND THE NORMATIVE OBJECTION

Daniele Bruno

THE ERROR THEORY is the view that normative judgments are beliefs that ascribe normative properties, but that these properties do not exist. According to the Error Theory, all of these judgments are therefore false.¹ One of the most formidable challenges to this theory is the Normative Objection. The idea behind this objection, most forcefully put forward by Ronald Dworkin and Thomas Nagel, is simple.² If we compare the plausibility of the Error Theory, in light of the most convincing arguments in favor of it, with the plausibility of some of our most deeply held normative judgments, we should come down firmly in favor of our normative judgments. To put the point differently: if we are to either believe that the Error Theory is true or that it is not true that we ought not to torture children for fun, then the only reasonable conclusion to draw is that the Error Theory is false.

In his book *Unbelievable Errors*, Bart Streumer has recently offered a novel and powerful defense of the Error Theory against this objection. He grants that our most deeply held normative judgments appear a lot more plausible than the Error Theory, to the extent that it seems bizarre to give them up in favor of it. But, he argues, this is not because the Error Theory is false. Instead, or so Streumer claims, it is because we *cannot believe* the Error Theory that it seems implausible when viewed against the background of our firmly held normative beliefs.

Though this *Unbelievability Thesis* is certainly highly controversial, I will not attempt to challenge it in what follows. Instead, I will argue that even if Streumer is correct in claiming that we cannot believe the Error Theory, this helps little to deflect the force of the Normative Objection. As I shall show, we can challenge the soundness of the main arguments that Streumer fields in support of the Error Theory through a kind of *Undermining Normative Objection* without appealing to

- 1 Important defenders of encompassing Error Theories include Olson (*Moral Error Theory*) and Streumer (*Unbelievable Errors*), though only the latter is explicit about it also extending to epistemic normativity.
- 2 See Dworkin, "Objectivity and Truth"; Nagel, *The Last Word*.

the plausibility of the theory as a whole. I shall proceed as follows. In section 1, I will lay out the Normative Objection and Streumer's unbelievability defense against it. In section 2, I will very briefly sketch Streumer's main arguments for the Error Theory. In section 3, I lay out the Undermining Normative Objection in detail. I conclude in section 4.

1. THE NORMATIVE OBJECTION AND THE UNBELIEVABILITY RESPONSE

Following Streumer's preferred reconstruction, we can understand the standard formulation of the Normative Objection along the following lines:

1. If a claim *C* and a philosophical theory *T* cannot both be true, and if *C* is much more plausible than *T*, we should reject *T*.
2. The claim that we ought not to torture children for fun and the Error Theory cannot both be true.
3. The claim that we ought not to torture children for fun is much more plausible than the Error Theory.

Therefore:

4. We should reject the Error Theory.³

Streumer grants premises 2 and 3, but denies premise 1. The crucial problem with the premise, according to him, is that it overlooks an alternative explanation for the greater plausibility of *T*, other than *T*'s most likely being false.⁴ Instead, our greater confidence in *C* could also be explained by *T*'s unbelievability. A theory that we cannot believe surely will not appear plausible to us, no matter whether it is false or not. All we can then safely assert is:

- 1*. If a claim *C* and a philosophical theory *T* cannot both be true, and if *C* is much more plausible than *T*, either it is the case that we should reject *T*, or *T* is unbelievable.

This, in turn, only allows us to infer the following, weaker conclusion:

- 4*. Either we should reject the Error Theory or the Error Theory is unbelievable.

As noted, I shall not challenge Streumer in his assumption that the Error Theory is unbelievable.⁵ This would then leave the Normative Objection without any

3 Streumer, *Unbelievable Errors*, 173–75.

4 Streumer, *Unbelievable Errors*, 176.

5 For pertinent criticism, see Olson, "On the Defensibility and Believability of Moral Error

force against Streumer's version of the Error Theory. The Normative Objection does not give us any reason to reject the Error Theory, since the theory's unbelievability provides a suitable alternative explanation of its plausibility deficit.

Streumer's response to the Normative Objection turns fundamentally on the Unbelievability Thesis. His argument for this thesis proceeds from what he takes to be necessary conditions for the attitude of belief, conditions that Streumer holds could not be met by individuals convinced by the arguments for the Error Theory.⁶ The Unbelievability Thesis thus only applies to the encompassing version of the Error Theory that Streumer himself defends—an Error Theory about all normative judgments, including judgments about reasons for belief.⁷ As Streumer himself admits, we are able to believe narrower kinds of error theory, as long as these do not extend to all judgments about reasons for belief.⁸ These type of theories, like the classical Moral Error Theory defended by Mackie and by Joyce, are thus unable to avoid the Normative Objection in Streumer's way.⁹

Alexander Hyun and Eric Sampson pick up on this fact, and try to show how it comes back to haunt Streumer in the end:

Although we cannot believe the Error Theory, we can come close to believing the Error Theory, and Streumer has argued that we have reason to do so. Streumer recognizes that a way to come close to believing the Error Theory is to believe those theses that are parts of the Error Theory, and surely Moral Error Theory is a part of the Error Theory. So, if there are reasons to come close to believing the Error Theory, then there are reasons to believe Moral Error Theory, and as a result our deepest and most important moral convictions are indeed threatened.¹⁰

I think Hyun and Sampson have the right hunch here. However, they do not quite manage to put their finger on the precise way in which the possibility of believing the Moral Error Theory causes problems for Streumer. Streumer himself makes this clear in response.

The arguments [for the Error Theory] will make us believe a Moral Error

Theory"; and Forcehimes and Talisse, "Belief and the Error Theory."

6 See Streumer, *Unbelievable Errors*, ch. 9.

7 I will refer to this encompassing version as "the Error Theory" in what follows.

8 Streumer, *Unbelievable Errors*, 152.

9 Mackie, *Ethics*; Joyce, *The Myth of Morality*.

10 Hyun and Sampson, "On Believing the Error Theory," 640.

Theory only if we mistakenly think that these arguments fail to apply to judgements about reasons for belief.¹¹

Though Streumer does not quite make it explicit, I think his reasoning is as follows. There are a number of true premises which together, via intermediary conclusions, form a sound argument for the Error Theory. Since the Error Theory is unbelievable, however, they do not succeed in making us believe it. Nonetheless, the arguments are able to make us believe the individual intermediary conclusions supported by their premises. These are the “parts of the Error Theory” that the Error Theorist’s arguments exert pressure on us to believe, since the relevant sub-arguments proceed only from believable premises to believable conclusions. Hyun and Sampson, however, seem to suggest that Streumer would be forced to believe a Moral Error Theory by reasoning like this:

1. The Error Theory is true.
2. If we cannot believe a true theory, we have reason to come close to believing it.
3. We can come close to believing the Error Theory by believing the Moral Error Theory.

Therefore:

4. We have reason to believe the Moral Error Theory.

Unlike the arguments supporting the intermediate conclusions, this case for believing the Moral Error Theory involves an unbelievable premise—the truth of the Error Theory. And since, as Streumer holds, arguments that turn on unbelievable propositions will not succeed in making us believe anything, he finds himself, *pace* Hyun and Sampson, under no pressure to come close to believing the Error Theory by believing a Moral Error Theory.¹²

However, there is a different, more efficient way to bring to bear the Normative Objection on the Error Theory via a detour through the Moral Error Theory. Instead of trying to give us reason to disbelieve the Error Theory as a whole, this objection systematically challenges the arguments in support of it. Before formulating it, however, I will have to briefly survey the general shape of the arguments that Streumer fields for the Error Theory.

¹¹ Streumer, *Unbelievable Errors*, 177.

¹² Furthermore, it is hard to see how an argument that relied on the truth of the Error Theory could ever entail that there is reason to believe the parts of the Error Theory that are believable, since, if the Error Theory were true, there would be no reason to believe anything.

2. STREUMER'S ARGUMENTS FOR THE ERROR THEORY

In chapters 2 to 7 of *Unbelievable Errors*, Streumer puts forward three main lines of argument in defense of the Error Theory. Here is Streumer summarizing their upshots.

The reduction argument shows that

1. If there are normative properties, these properties are identical to descriptive properties,

and the false guarantee and regress objections show that

2. If there are normative properties, these properties are not identical to descriptive properties.

These claims together entail that normative properties do not exist. But the symmetry objection shows that

3. Normative judgements are beliefs that ascribe normative properties.

These three claims together entail that the Error Theory is true: they together entail that normative judgements are beliefs that ascribe normative properties, but that these properties do not exist.¹³

Streumer's endorsement of the Unbelievability Thesis lends these three lines of argument a slightly peculiar standing. As Streumer puts it, he believes they are "*sound* arguments that together *seem* to show that the Error Theory is true."¹⁴ He hastens to clarify that this does not mean that the arguments are only seemingly sound. Nonetheless, they cannot lead us to believe that the Error Theory is true, because we can follow an argument to its conclusion only if that conclusion is believable. Nonetheless, the fact that the arguments *are* sound is crucial. Were there no sound arguments that together entailed the truth of the Error Theory, the theory would not only fail to be interesting, but we would most likely have reasons for believing that it is false, given its strongly counterintuitive implications.

3. THE UNDERMINING NORMATIVE OBJECTION

The soundness of the arguments just surveyed is what is challenged by the Un-

13 Streumer, *Unbelievable Errors*, 103–4.

14 Streumer, *Unbelievable Errors*, 153.

dermining Normative Objection. To present this challenge, one need not consider the details of the arguments. What suffices, rather, is their general form. Here, then, is my argument:

1. The main argument for the Error Theory has the following general form:
 - a. All members of set S are x , if they exist.¹⁵
 - b. All members of set S are non- x , if they exist.
 - c. No members of set S exist.
2. If an argument that has the aforementioned general form is sound, then there is a derivative argument with the following form that is also sound:
 - a'. All members of subset S_1 are x , if they exist.
 - b'. All members of subset S_1 are non- x , if they exist.
 - c'. No members of subset S_1 exist.
3. If the main argument for the Error Theory is sound, then there is a derivative argument for the Moral Error Theory that is also sound. (From 1 and 2)
4. If there is a sound argument for the Moral Error Theory, then we should believe the Moral Error Theory.
5. If a claim C and a philosophical theory T cannot both be true, and if C is much more plausible than T , either it is the case that we should reject T , or T is unbelievable.
6. If the Moral Error Theory is true, then it cannot be true that we ought not to torture children for fun.
7. The claim that we ought not to torture children for fun is much more plausible than the Moral Error Theory.
8. We can believe the Moral Error Theory.
9. We should reject the Moral Error Theory. (From 5, 6, 7, and 8)
10. The derivative argument for the Moral Error Theory is not sound. (From 4 and 9)
11. Therefore, the argument for the Error Theory is not sound. (From 3 and 10)

This argument is clearly valid, since it employs only a simple succession of straightforward applications of *modus ponens* and *modus tollens*. Let me thus offer a few words on its premises. I believe premise 1 involves a fair reproduction of the general structure of Streumer's argument as laid out above. Premise

15 S here being the set composed of all normative properties, and x being [identical to a descriptive property].

2 appears impossible to deny, on pain of going against elementary logic. If we can disprove the existence of all members of a set by pointing to contradictory qualities all such members would have to have, then we can equally do so for members of a subset, since these necessarily share the same qualities. Applied to the case for the Error Theory, S_1 would then of course be the set of moral properties. Since these do indeed form a subset of the broader category of normative properties, the first intermediate conclusion, 3, follows from 1 and 2.¹⁶

Premises 6, 7, and 8 are, as I have already mentioned above, all explicitly endorsed, or at least not rejected, in various places of Streumer's discussion of the Normative Objection. This leaves premises 4 and 5 as the only potential points of defense for the defender of the Error Theory.

One move that might immediately come to mind would be to qualify premise 4 in the same way that we qualified the first premise of the original Normative Objection above. We might say that there being a sound argument for the Moral Error Theory does not show that we should accept the Moral Error Theory, but merely that either it is so, or the theory is unbelievable. This however does not help the Error Theorist in the current situation. As Streumer acknowledges, it is not the case that the Moral Error Theory, one of the premises of the argument, or a combination of them is unbelievable. Even though the argument is structurally isomorphic, the premises of the argument proposed in 2 are independent of those involved in the case for the Error Theory. There is thus a set of true premises, all of which can be believed, which together entail a conclusion that can equally be believed.

The situation is rendered slightly more complicated by the fact that the derivative argument in premise 2 may not, after all, succeed in showing that people like Streumer himself, who also believe the premises of the argument for the Error Theory, should accept the Moral Error Theory. This is because these people may not be able to assess premises a' , b' , and c' independently from a , b , and c . Therefore, they may not be able to follow the argument to its conclusion, since doing so would amount to (also) accepting the unbelievable Error Theory.

16 A caveat: I have been simplifying matters a bit by speaking of the *main argument* for the Error Theory. I have been hedging in this way because showing that normative properties do not exist is obviously not sufficient for a case for the Error Theory. One also needs to show (as Streumer attempts to) that normative judgments do actually ascribe these properties, in order to forestall non-cognitivist alternatives. This complication does not matter for my purposes here, however. If normative discourse on the whole is cognitive, then moral discourse, as a particular type of normative discourse, surely is as well. We can therefore safely assert, with 3, that if the argument for the Error Theory is sound, then there is a derivative argument for the Moral Error Theory that is also sound.

Premise 4 thus may not be true as long as the “we” refers to those already convinced of Streumer’s arguments.

But this fact does not rob the Undermining Normative Objection of its dialectical force. The point of the objection is not to show that actual defenders of the Error Theory such as Streumer are committed to believing implausible conclusions. Rather, it is to show that there is a problem with Streumer’s argument for the Error Theory. His argument entails the soundness of the derivative argument for the Moral Error Theory, and the unbelievability defense simply is not available for this latter argument. Streumer owes us a different kind of explanation of what is wrong with this argument *qua* argument, irrespective of whether *he* himself can ultimately believe it.¹⁷ If such an explanation cannot be provided, then we are forced to conclude that Streumer’s arguments for the Error Theory cannot be sound either, at least if we accept the presuppositions of the Normative Objection (i.e., 6, 7, and 8).¹⁸

4. CONCLUSION

As I have shown, Streumer’s attempt to deflect the Normative Objection by appeal to the unknowability of the Error Theory fails. This is an important result. I take it that one of the most interesting features of Streumer’s project of joining a case for the Error Theory with a defense of the Unbelievability Thesis is that it allows him to put forward his arguments for the Error Theory from a dialectically much more comfortable position. The Unbelievability Thesis, or so Streumer seems to think, allows us to engage in a carefree pursuit of a case for the Error Theory in its most radical form, without having to engage in many of the most

17 As an anonymous reviewer for *JESP* points out, a more serious problem for the Undermining Normative Objection would result if *nobody* was able to entertain *a'*, *b'*, and *c'* without thereby coming to believe *a*, *b*, and *c*. It is true that if it were so, the defender of the Error Theory would have an out here. However, it seems to me that the underlying assumption does not stand up to scrutiny. Even though it is entailed by the original argument for the Error Theory, the derivative argument can solidly stand for itself: its premises *a'*, *b'*, and *c'* can be understood and justified without reference to *a*, *b*, and *c*. To make this clear by filling in the variables: a person can be convinced that moral properties, if they really existed, would have to be both identical to descriptive properties and not identical to descriptive properties, without thereby being led to believe anything about nonmoral (in particular: epistemic) normativity. This possibility is all that is required for the undermining normative objection to retain its force.

18 The Undermining Normative Objection of course does not tell us anything about *where* Streumer’s arguments for the Error Theory go wrong, just that there has to be a flaw in them somewhere. For a recent attempt to identify problems in Streumer’s argument, see Laskowski, “Reductivism, Nonreductivism, and Incredulity about Streumer’s Error Theory.”

pressing objections traditionally fielded against it. If the Undermining Normative Objection as laid out above is sound, however, then Error Theorists such as Streumer cannot sit back and relax quite as early. At least one of the traditional objections against the Error Theory, and quite possibly the most formidable one, still is very much on the table even if we accept the Unbelievability Thesis.¹⁹

Humboldt University of Berlin
daniele.bruno@hu-berlin.de

REFERENCES

- Dworkin, Ronald. "Objectivity and Truth: You'd Better Believe It." *Philosophy and Public Affairs* 25, no. 2 (Spring 1996): 87–139.
- Forchimes, Andrew T., and Robert B. Talisse. "Belief and the Error Theory." *Ethical Theory and Moral Practice* 19, no. 4 (February 2016): 849–56.
- Hyun, Alexander, and Eric Sampson. "On Believing the Error Theory." *Journal of Philosophy* 111, no. 11 (November 2014): 631–40.
- Joyce, Richard. *The Myth of Morality*. Cambridge: Cambridge University Press, 2001.
- Laskowski, N. G. "Reductivism, Nonreductivism, and Incredulity about Streumer's Error Theory." *Analysis* 78, no. 4 (October 2018): 766–76.
- Mackie, J.L. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin, 1977.
- Nagel, Thomas. *The Last Word*. New York: Oxford University Press, 1997.
- Olson, Jonas. *Moral Error Theory*. Oxford: Oxford University Press, 2014.
- . "On the Defensibility and Believability of Moral Error Theory: Reply to Evers, Streumer, and Toppinen." *Journal of Moral Philosophy* 13, no. 4 (June 2016): 461–73.
- Streumer, Bart. *Unbelievable Errors*. Oxford: Oxford University Press, 2017.

19 I would like to thank David Heering, Bart Streumer, and two anonymous referees for *JESP* for their helpful comments on earlier versions of this article. Work on this project has been supported by the Einstein Stiftung Berlin.

DOES INITIAL APPROPRIATION CREATE NEW OBLIGATIONS?

Jesse Spafford

A SIGNATURE CLAIM of entitlement theories of justice is that people have private property rights over objects. Additionally, proponents of these theories generally maintain that these rights can be established *unilaterally*: by performing the right kind of appropriative act, an individual can convert unowned natural resources into private property without having to obtain the consent of others. However, many philosophers have objected to this latter claim as follows:

1. Morally equal people do not have the power to unilaterally impose obligations on one another (i.e., impose such obligations without consent).
2. The power to unilaterally appropriate is a power to unilaterally impose obligations on others, as they are now obligated to refrain from using the appropriated thing.
3. Thus, people lack the power to unilaterally appropriate.

This *moral equality argument*—or some variant thereof—has been advanced by a number of philosophers, typically as part of a broader account of the conditions under which such appropriation would be possible.¹ However, a recent argument advanced by Bas van der Vossen threatens the second premise, as it raises the possibility that initial appropriation does not *create* new obligations for others, but rather *alters* the requirements implied by their already existing obligations in a way that is unproblematic vis-à-vis moral equality.² This paper

- 1 Bas van der Vossen helpfully catalogs a number of such proponents—primarily Kantians—including Gibbard (“Natural Property Rights”), Ripstein (*Force and Freedom*, 272), and Waldron (“The Right to Private Property,” 265–67, 280, and “Kant’s Legal Positivism,” 1557). He also identifies the many philosophers who have read Kant as making this argument, including Flikschuh (*Kant and Modern Political Philosophy*, 136, 141, 228) and Stilz (*Liberal Loyalty*, 45, 55). A similar argument is made by Wenar, though he does not lean as heavily on the moral equality premise (“Against Moral Responsibility,” 806–7).
- 2 Van der Vossen, “Imposing Duties and Original Appropriation.”

attempts to rescue the moral equality argument from van der Vossen's objection by showing that acts of initial appropriation do, in fact, imply a morally problematic power to impose duties on others.³

1. THE DUTY ALTERATION OBJECTION

To begin, it will be helpful to introduce the normative ontology that van der Vossen employs when objecting to the second premise. Specifically, he draws a distinction between *obligations* and their associated *requirements*, where the former are expressed by general normative propositions—e.g., “Q is obligated not to touch P's body without P's permission”—while the latter are expressed by action-specific normative propositions whose truth values are a function of both certain obligations and certain facts about the world. For example, if (a) Q has an obligation not to touch P's body and (b) P has hair, then Q has the requirement that she not touch P's hair.

Given that requirements are a function of both obligations and empirical facts, it follows that there are two ways that one might modify those requirements, namely, by changing some normative fact(s) about which obligations obtain or by changing the relevant empirical facts.⁴ What van der Vossen calls *duty-creation*, then, is the Hohfeldian power to change some person's requirements by generating new obligations for her (or, perhaps, by changing the content of her existing obligations); by contrast, *duty-alteration* is the power to change requirements by changing relevant empirical facts.⁵ Thus, when P grows out her hair, she is exercising her power of duty-alteration rather than duty-creation, as she changes Q's requirements without changing Q's obligations.

With this distinction in place, premise 1 of the argument from moral equality—which denies that people have the power to impose obligations on one another—can now be understood as a denial that people have the power of duty-creation. And van der Vossen is happy to concede this point, as his view is that no one has such a power. However, this view also leads him to reject arguments that cite the unacceptability of duty-creation as the reason that a particular mor-

- 3 The paper will also function as a reply to related defenses of initial appropriation, including those made by Gaus and Lomasky (“Are Property Rights Problematic?”) and Simmons (“Original-Acquisition Justifications of Private Property”).
- 4 Van der Vossen, “Imposing Duties and Original Appropriation,” 69.
- 5 Van der Vossen, “Imposing Duties and Original Appropriation,” 70. Technically, van der Vossen says the particular *acts* are instances of either duty-creation or duty-alteration (as opposed to the *power* to carry out such acts as just asserted). However, using the terms to refer to powers helps to clarify the discussion, with the slight misattribution then being necessary for introducing the term as it is to be used later in the text.

al power does not obtain. Indeed, while many have objected to initial appropriation on the grounds that such appropriation imposes new obligations upon others, van der Vossen argues that initial appropriation is merely an exercise of the power of duty-alteration, as he posits that people have the following *natural conditional right to appropriate*:

For all persons P and Q at times t_1 and t_2 , P has a right against Q that [Q respect P as the rightful owner of O at t_2 on the condition that P performs appropriative act A on object O under conditions C at t_1].⁶

The right that P has in this case is that the bracketed conditional obtains—with Q then having a correlative obligation to make the bracketed conditional obtain. More specifically, this obligation entails that Q act in the way specified by the consequent (respect P as the rightful owner) when the antecedent obtains (P performs A on O). Then, just as P growing out her hair changes some of the requirements of Q 's obligation not to touch P —but does so without adding to or modifying Q 's existing obligations— P doing A changes the requirements of Q 's natural conditional obligation via the alteration of an empirical fact.

Given that P 's appropriation merely changes Q 's requirements, the act would not entail the prohibited power to impose obligations. In other words, there are two sorts of moral powers, one problematic (duty-creation), one unproblematic (duty-alteration), with both hair growing and initial appropriation implying only the latter. Thus, to rescue the moral equality argument, one must show that there is some other principled basis for demarcating hair growing from initial appropriation such that it can be maintained that the latter is uniquely problematic. It is this task that the paper will take up in the next section.

2. THE REVISED MORAL EQUALITY ARGUMENT

The central contention of the duty-alteration objection is that, while both hair growing and initial appropriation change others' requirements, neither imposes new obligations. However, this section will argue that, in fact, the P who grows out her hair changes *neither* Q 's obligations *nor* Q 's requirements. By contrast, the P who appropriates some unowned thing *does* change Q 's requirements. Thus, the moral equality argument can sidestep the duty-alteration objection by maintaining that it is the power to *unilaterally impose new requirements* that runs contrary to the assumption of human moral equality.

To begin, note that a foundational assumption of the duty-alteration objec-

6 Van der Vossen, "Imposing Duties and Original Appropriation," 74. While brackets appear in the cited text, some variables have been replaced for stylistic consistency.

tion is that if *P* grows out her hair, then *Q* has a new requirement not to touch *P*'s hair. But why think, as van der Vossen does, that *Q* lacked this requirement when *P* lacked hair? This crucial premise would seem to rest on the following tacit argument:

1. *P* does not have hair.
2. If *P* does not have hair, then *Q* cannot touch *P*'s hair.
3. *Q* is required to ϕ only if she can ϕ (OIC: "ought" implies "can").
4. Thus, when *P* has no hair, it is false that *Q* is required not to touch *P*'s hair (i.e., there is no requirement that *Q* not touch *P*'s hair).

The problem with this argument, of course, is that it is invalid. To see this, note that OIC together with the proposition that *Q* cannot touch *P*'s hair implies that it is not required that *Q* touch *P*'s hair—not the asserted conclusion that *Q* is *not forbidden* from touching *P*'s hair (where *Q* is *forbidden* from ϕ -ing just in case she is required not to ϕ). To reach *this* conclusion via OIC, it would have to be the case that *Q* is unable to *avoid* touching *P*'s hair; however, in the stipulated case where *P* lacks hair, *Q* is fully able to avoid touching *P*'s hair. Thus, the conclusion that *P* is not forbidden from touching *P*'s (nonexistent) hair does not follow from the argument's premises.

Given the failure of this argument, the proponent of the moral equality argument could insist that, just as *Q* is forbidden from touching *P*'s hair when *P* has hair, *Q* is equally forbidden from touching *P*'s (nonexistent) hair when *P* does *not* have hair—and, thus, that *P* does not impose any new requirement on *Q* by growing her hair. By contrast, the *P* who appropriates some object *O* *does* impose new requirements upon *Q*, as *Q* was free to use *O* prior to its appropriation but becomes forbidden from such use as soon as *P* appropriates *O*. Thus, there is an important difference between hair growing and initial appropriation: the latter imposes novel requirements on others while the former does not.

This difference, in turn, allows for a restatement of the moral equality argument where what is proscribed is not the power to unilaterally impose novel obligations but, rather, the power to unilaterally impose novel *requirements*. While there is no contradiction in morally equal people having the power to unilaterally grow out their hair, there is a contradiction between people being moral equals and their having the power to unilaterally impose novel requirements on one another—or at least so the proponent of the moral equality argument might maintain. Thus, she would be able to sidestep the duty-alteration objection, as she can insist that, *contra* van der Vossen's claim, there is an important moral difference between initial appropriation and pedestrian activities like hair growing, even while conceding his contention that neither imposes novel obligations.

3. THE "OUGHT NOT" IMPLIES "CAN" OBJECTION

In response to this proposal, three objections might be raised. First, it might be objected that there is an easy fix for the invalid argument presented in the previous section: simply replace premise 3 (Q is required to ϕ only if she can ϕ) with the premise that Q is *forbidden* from ϕ -ing only if she can ϕ . Given this premise, Q would not be forbidden from touching P 's hair when P has no hair, with this action only becoming forbidden when P grows out her hair. Thus, hair growing would impose novel requirements on others, collapsing the proposed distinction between hair growing and appropriation.

However, while this replacement would render the argument valid, such a move would come at the expense of the plausibility of the third premise. Note that OIC is already controversial, with many arguments having been raised against it.⁷ However, even if one concedes that "ought" implies "can," there is little reason for thinking that "ought *not*" implies "can" (ONIC) as the amended premise 3 contends. Indeed, a quick survey of the prominent arguments for OIC reveals that none of the posited reasons for endorsing OIC can be appealed to in support of ONIC.⁸ For example, David Copp argues that a moral theory that required a person to ϕ when she cannot ϕ would be unfair—but moral theories cannot be unfair in this way.⁹ Thus, he concludes that one is required to ϕ only if one can ϕ . However, even if one grants this argument, there is nothing seemingly unfair about a moral theory that *forbids* a person from doing something she cannot do. Given the absence of such unfairness, it would then follow that Copp's argument for OIC cannot be repurposed to support ONIC.

Another popular line of argument for OIC is that this principle is needed to explain a number of facts about moral reasoning. For example, Frances Howard-Snyder argues that, if OIC were false, then we could not adequately explain (a) why an agent who cannot ϕ (where she otherwise ought to ϕ) ought to do the "second-best" thing instead, (b) why an agent who ought to ϕ also ought to ψ when ψ -ing is a necessary condition of her ϕ -ing, and (c) why there are *prima*

7 For some prominent arguments against this premise see Sinnott-Armstrong, "Ought' Conversationally Implies 'Can'"; Saka, "Ought Does Not Imply Can"; Ryan, "Doxastic Compatibilism and the Ethics of Belief"; Graham, "Ought' and Ability"; Waller, "Against Moral Responsibility," 179–89; and Talbot, "The Best Argument for 'Ought Implies Can' Is a Better Argument against 'Ought Implies Can.'"

8 These arguments all use slightly different moral language, with some being concerned with "ought" while others are concerned with obligations (typically "all things considered," *ultima facie* obligations). However, these generally seem to correspond to the way "requirement" is being used in this paper.

9 Copp, "Ought' Implies 'Can,' Blameworthiness, and the Principle of Alternate Possibilities."

facie obligations that are *overridden* by other obligations (as opposed to the agent simply having both obligations simultaneously).¹⁰ In each case, Howard-Snyder contends that the best explanation is that an agent only ought to ϕ if she can ϕ . However, again, one can fully concede this point while still denying ONIC: even if it is true that OIC must be true to explain (a), (b), and (c), ONIC does not appear to do any important explanatory work. Thus, this argument for OIC also cannot be generalized to defend ONIC.

Finally, there is a popular strategy for defending OIC that appeals to reasons for actions. For example, Peter Vranas argues that (1) an agent has an obligation to ϕ only if she has reason to ϕ , (2) she has reason to ϕ only if ϕ -ing is a potential option for her, and (3) ϕ -ing is a potential option for her only if she can ϕ ; thus, she has an obligation to ϕ only if she can ϕ .¹¹ Similarly, Bart Streumer argues that there cannot be a reason for an agent to ϕ if she lacks the ability to ϕ , and thus, she cannot have the *most reason* to ϕ if she is unable to ϕ .¹² He contends that this latter claim has the same truth conditions as the claim that it cannot be the case that a person ought to ϕ when she is unable to ϕ (i.e., OIC); thus, OIC is true.

Again, it does not appear that this argument can be repurposed to support ONIC. While it may be true that one must be able to ϕ if one is to have a reason to ϕ —with reason to ϕ being a necessary condition of being required to ϕ —it is not obvious that one must be able to ϕ to have reason *not* to ϕ . To see this, consider some of the supporting arguments Streumer gives for thinking that one must be able to ϕ if one is to have reason to ϕ . First, he argues that if one could have reason to ϕ without having the ability to ϕ , one could have “crazy” reasons like a reason to jump thirty-thousand feet into the air to stop a plane from crashing.¹³ However, given that it is absurd to think we have such reasons, he contends that reasons are ability constrained. However, being *forbidden* from doing things one cannot do generates no such “crazy” reasons, as there seems to be nothing “crazy” about having reason *not* to jump thirty-thousand feet into the air, for example.

Alternatively, Streumer argues that if agents can have reasons to do things they cannot do (i.e., OIC is false), then it will turn out that they will have most reason to do what they cannot (e.g., go back in time and stop all the wars).¹⁴ Thus, they will have to try to spend their lives pointlessly trying to do the impossible—an absurd conclusion that requires the rejection of the premise that

10 Howard-Snyder, “‘Cannot’ Implies ‘Not Ought,’” 236–41.

11 Vranas, “I Ought, Therefore I Can,” 171–72.

12 Streumer, “Reasons and Impossibility,” 351–58.

13 Streumer, “Reasons and Impossibility,” 358–59.

14 Streumer, “Reasons and Impossibility,” 365.

OIC is false.¹⁵ Again, though, this argument cannot be repurposed to establish ONIC, as being forbidden from carrying out undoable actions will not demand any sort of pointless efforts on the part of agents. Thus, arguments for OIC that appeal to the relationship between ability and reasons cannot be employed to defend ONIC.

There is one argument for OIC that *can* be repurposed to defend ONIC. This argument contends that the point of moral evaluation of actions is to aid in deliberation; however, given that undoable actions are not objects of deliberation, it is inappropriate to assign the moral predicate of “required” to such actions.¹⁶ Thus, one is only required to do actions that one can do (i.e., OIC)—a conclusion that can be generalized to defend ONIC if one insists that it is inappropriate to assign *any* moral predicates to actions that are not the objects of deliberation (i.e., undoable actions can be *neither* required nor forbidden).

The problem with this argument is its contention that moral predicates can only be applied to those actions that factor into deliberation. Suppose that some agent is a pacifist such that she would not even consider harming another person. For such a person, the possible action of killing her friend would never enter into her deliberation process. Thus, if one takes moral predication to be constrained by deliberation, it would then follow that the pacifist is not required to refrain from murdering her friend. However, this would seem to be a *reductio* of the premise that moral predicates only apply to those actions that factor into deliberations.¹⁷ Given that this *reductio* undermines the only apparent argument for ONIC, there would seem to be no basis for rejecting the revised moral equality argument’s contention that *Q* is forbidden from touching *P*’s hair even when *P* lacks hair.

4. THE OCCUPATION OBJECTION

There is a second possible objection that, like the first, contends that both initial appropriation and hair growing impose new requirements. Suppose that *P* grows out her hair such that it comes to occupy space *S* at time *T*. This objec-

15 Streumer, “Reasons and Impossibility,” 365. Streumer also has an additional argument, but I think it ultimately collapses into the argument from crazy reasons, though I will not argue for that here.

16 Copp makes a point along these lines (“‘Ought’ Implies ‘Can,’ Blameworthiness, and the Principle of Alternate Possibilities,” 273–74).

17 Peter Graham makes a related point when arguing that we need not posit that “ought implies can” when explaining our process of deliberation. He notes that we not only exclude undoable actions at the outset of our deliberation, but also actions we have no intention of doing (“‘Ought’ and Ability,” 371–72).

tion argues that it would now be true that *Q* is required not to move her hand through *S* at *T*—something that was not true before *P* grew out her hair. Thus, *contra* the argument of section 3, hair growing *does* result in requirement change.

In response to this objection, recall that moral requirements are specific to particular actions—i.e., the propositions that express such requirements will refer to some particular action and attach a moral predicate (e.g., “required” or “forbidden”) to a description of that action. Further, note that actions are spatio-temporal entities that include particular bits of matter moving through particular spaces at particular times. Thus, actions are properly individuated on the basis of the physical parts that compose them. For example, consider the case of *P* walking through an unoccupied doorway versus *P* passing through that same doorway but trampling *Q* in the process. Even though both might fall under the description of “*P* moves through the doorway,” these actions are properly understood as distinct because (a) *P*’s physical motion will differ across the two cases (in one she shoves *Q* as part of her movement) and (b) one action includes *Q* while the other does not.

Given that requirements are specific to particular actions, it becomes clear that the occupation objection is wrong to claim that the action of *Q* moving her hand through *S* at *T* becomes newly forbidden when *P*’s hair comes to occupy *S*. Indeed, this claim only seems true because the hand-moving action is under-described. If one specifies that the action in question is *Q* moving her hand through *S* and brushing *P*’s hair in the process, then the action would still be forbidden even before *P* grows out her hair. By contrast, if the action in question is *Q* moving her hand through unoccupied space *S*, then the action would not be forbidden before or after *P* grows out her hair. Rather, it would merely become *impossible* for *Q* to do this action without the deontic status of the action changing. Thus, the occupation objection rests on a false premise and the revised moral equality argument can be sustained.¹⁸

5. THE ETERNALISM OBJECTION

While the first two objections maintained that *both* hair growing and initial appropriation impose new requirements on others, the third objection contends that *neither* action imposes such requirements. Specifically this objection holds

18 Another way of putting this point is that, even if the proposition “*Q* is forbidden from moving her hand through *S* at *T*” changes truth values when *P* grows out her hair, this proposition does not express a *requirement* because it is not specific to a particular action. For a new requirement to be imposed, some proposition that expresses a requirement must change from false to true—a change that appropriation brings about but hair growing does not.

that a novel requirement is imposed just in case the truth value of the proposition that expresses that requirement changes from false to true. However, there is a popular metaphysical view that denies that propositions can have different truth values at different times. Thus, the distinction between initial appropriation and hair growing cannot be stated in terms of requirement change, as requirements cannot change across time.

This objection assumes the popular position that propositions are *eternal* in the sense that their truth value does not change over time, thanks in part to the fact that their content is temporally indexed.¹⁹ For example, when a person in Chicago says, “It is raining” at 12:45 PM on December 20, 1989, she is really expressing the proposition that it is raining in Chicago at 12:45 PM on December 20, 1989. Further, the eternalist claims that this proposition is true both at the time the sentence is uttered and thirty years later when the speaker is in New York and the weather is clear both there and in Chicago. Thus, in contrast to *temporalist* views, which hold that the sentence “It is raining” expresses a proposition that is true when it is raining and false when it is not—i.e., a proposition that changes truth values across time—the eternalist position holds that propositions are timelessly true (or timelessly false).

Given an eternalist view of propositions, it would then follow that the truth values of requirement-expressing propositions do not change across time—i.e., no new requirements could be imposed. To see this, consider the case of an object *O* that is unowned at t_1 but is appropriated by *P* at t_2 and then owned by *P* at time t_3 . Given the general obligation not to touch others’ property without their consent coupled with the empirical fact that *P* carried out an appropriative act at t_2 , the following proposition would be true: *Q* is required not to touch *O* at t_3 . However, if this proposition is true, then, on the eternalist view, it is equally true at t_1 when *O* is unowned as it is at t_3 when *P* has come to own it. Thus, *contra* the revised moral equality argument, *P*’s appropriative act does not generate a new requirement that *Q* not touch *O* at t_3 .

The most straightforward reply to this objection is to simply concede the point and restate the revised moral equality argument in terms that are compatible with eternalism. This twice-revised argument would avoid talk of requirement change and, instead, put things in terms of the *counterfactual requirement differences* associated with initial appropriation versus hair growing. Specifically, note that when *P* appropriates object *O*, *Q* ends up with a requirement that she would not have had if *P* had not appropriated *O*, namely, the requirement not to touch *O*. By contrast, *Q* is equally required not to touch *P*’s hair in the world

19 While this is the dominant view, there are dissenters, including Prior, “Thank Goodness That’s Over”; and Brogaard, *Transient Truths*.

where *P* grows out her hair and in the world in which she keeps her head shaved. Indeed, the arguments above have sought to demonstrate that *P*'s lack of hair does nothing to negate the proposition that *Q* is forbidden from touching *P*'s hair. Thus, even granting the eternalist claim about propositions, there is still an important difference between hair growing and initial appropriation: the latter generates a counterfactual requirement difference while the former does not. It is this power—the power to unilaterally burden people with requirements they would not have otherwise had—that is incompatible with moral equality. Or so the proponent of the moral equality argument could maintain in the face of the duty-alteration and eternalism objections.

6. CONCLUSION

This paper has attempted to rescue the moral equality argument from the objection that the power to appropriate is no different from the power to grow out one's hair. Specifically, it has attempted to demonstrate that only initial appropriation imposes new requirements on others, making the power to unilaterally appropriate morally problematic in a way that the unilateral power to grow out one's hair is not. Finally, the paper rejected three additional reasons for thinking that initial appropriation and hair growing are of a kind (e.g., because both/neither impose new requirements on others). Thus, it concludes that the moral equality argument survives the duty-alteration objection.²⁰

The Graduate Center, CUNY
jesse.e.spafford@gmail.com

REFERENCES

- Brogaard, Berit. *Transient Truths: An Essay in the Metaphysics of Propositions*. Oxford: Oxford University Press, 2012.
- Copp, David. "'Ought' Implies 'Can,' Blameworthiness, and the Principle of Alternate Possibilities." In *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, edited by David Widerker and Michael McKenna. Burlington, VT: Ashgate Press, 2003.

20 I am indebted to Bas van der Vossen for his sustained discussion of this subject with me and his many helpful comments on earlier drafts of this paper. I am also grateful to the American Council of Learned Societies for providing me with a Mellon/ACLS Dissertation Completion Fellowship that helped make this research possible.

- Flikschuh, Katrin. *Kant and Modern Political Philosophy*. Cambridge: Cambridge University Press, 2000.
- Gaus, Gerald F., and Loren E. Lomasky. "Are Property Rights Problematic?" *The Monist* 73, no. 4 (October 1990): 483–503.
- Gibbard, Alan. "Natural Property Rights." *Noûs* 10, no. 1 (March 1976): 77–86.
- Graham, Peter. "'Ought' and Ability." *Philosophical Review* 120, no. 3 (July 2011): 337–82.
- Howard-Snyder, Frances. "'Cannot' Implies 'Not Ought.'" *Philosophical Studies* 130, no. 2 (August 2006): 233–46.
- Prior, Arthur. "Thank Goodness That's Over." *Philosophy* 34, no. 128 (1959): 12–17.
- Ripstein, Arthur. *Force and Freedom: Kant's Legal and Political Philosophy*. Cambridge, MA: Harvard University Press, 2009.
- Ryan, Sharon. "Doxastic Compatibilism and the Ethics of Belief." *Philosophical Studies* 114, nos. 1–2 (May 2003): 47–79.
- Saka, Paul. "Ought Does Not Imply Can." *American Philosophical Quarterly* 37, no. 2 (April 2000): 93–105.
- Simmons, A. John. "Original-Acquisition Justifications of Private Property." *Social Philosophy and Policy* 11, no. 2 (Summer 1994): 63–84.
- Sinnott-Armstrong, Walter. "'Ought' Conversationally Implies 'Can.'" *Philosophical Review* 93, no. 2 (April 1984): 249–61.
- Stilz, Anna. *Liberal Loyalty*. Princeton, NJ: Princeton University Press, 2009.
- Streumer, Bart. "Reasons and Impossibility." *Philosophical Studies* 136, no. 3 (December 2007): 351–84.
- Talbot, Brian. "The Best Argument for 'Ought Implies Can' Is a Better Argument against 'Ought Implies Can.'" *Ergo* 3, no. 14 (2016): 377–402.
- van der Vossen, Bas. "Imposing Duties and Original Appropriation." *Journal of Political Philosophy* 23, no. 1 (March 2015): 64–85.
- Vranas, Peter B. M. "I Ought, Therefore I Can." *Philosophical Studies* 136, no. 2 (November 2007): 167–216.
- Waldron, Jeremy. "Kant's Legal Positivism." *Harvard Law Review* 109, no. 7 (1996): 1553–66.
- . *The Right to Private Property*. Oxford: Clarendon Press, 1988.
- Waller, Bruce. *Against Moral Responsibility*. Cambridge, MA: MIT Press, 2011.
- Wenar, Leif. "Original Acquisition of Private Property." *Mind* 107, no. 428 (October 1998): 799–817.