Philosophers' Imprint

VOLUME 24, NO. 11
SEPTEMBER 2024

Theories, Facts, and Meanings in Political Philosophy

Guido Pincione & Gregory Robson

University of Arizona *University of Notre Dame*

DOI: 10.3998/phimp.1909

Normative political theories can help us to articulate principles and values of political morality, but their value goes beyond that function. Voters, politicians, and other political actors frequently appeal to principles or values in advocating public policy, or take them into consideration in political decision-making. Of course, political actors may not always know a theory's specific details, but they can invoke the theory's principles or values, or approximations thereof. Now, the policies implemented or supported on the basis of those principles or values may be counterproductive or even independently objectionable. It is not difficult to imagine, for example, policymakers publicly appealing to democratic values in order to implement policies that create poverty or inequalities that no attractive political morality could condone. Our focal question, therefore, is the following: should the *predictable use* of a theory — not just its content and entailments — determine, at least in part, its theoretical value? More generally, if the principles or values of a normative political theory will be put to bad use, can the theory be bad *on that basis* — even if, as is often the case, few or no politically influential people will claim to know, or indeed know, the theory behind those principles or values?

In general, the consequences of *correctly* implementing a theory arguably bear on its acceptability.[1] A theory of justice whose correct implementation permitted chattel slavery would be highly implausible. The claim we defend, here, however, concerns *incorrect* implementation. We argue that theories that will predictably be put to bad use deserve harsher assessments than those that will predictably be put to better use. Theories that key political actors will predictably invoke to justify bad policy recommendations are thus bad theories, even when those recommendations are not logical consequences of those theories (even in conjunction with well-established factual propositions).

Standard discussions of political theories ignore these defects when assessing a theory's principles and values. "Misapplications" of a theory, so the thought goes, should not count against it. To take one example, suppose legislators will predictably use John Rawls's

1. This position is a corollary of the reflective-equilibrium methodology embraced, explicitly or implicitly, by many political theorists.

difference principle (which "comes into play at the stage of the legislature") to support policies aggravating poverty.[2] While the principle "dictates that social and economic policies be aimed at maximizing the long-term expectations of the least advantaged," the legislators instead invoke it to support policies that frustrate that goal, worsening the situation of the least well-off.[3] We are not aware of anyone arguing that such counterproductive uses of the difference principle would count against Rawls's theory. And we will see in sections II and III why any such critique may appear irrelevant. We shall argue, however, that correctly assessing a normative political theory requires one to consider political actors' predictable (mis)interpretations. We include here misinterpretations of the principles or values the theory defends, even if those misinterpretations make no claims about the relationship between the principles, values, and theory — after all, political actors seldom cite political theories. Our central thesis, therefore, is the following: The fact that a political theory will predictably be applied in ways that yield counterproductive or independently objectionable outcomes is a major reason to reject that theory. This is true even if the theory's meaning — more specifically, what we shall call its "narrow meaning," in conjunction with suitable empirical propositions — does *not* require any such applications.

I. Narrow and Broad Meanings of a Political Theory

By a "normative political theory," or "political theory," for short, we mean an articulated argument in favor of (i) important rules, principles, or values of political morality (the difference principle, equality, liberty, etc.) or (ii) important political institutions (the state, democracy, private property, the rule of law, etc.). For present purposes, a rule, institution, etc. counts as important just in case its existence or promotion has significant implications for the lives of many of a polity's members. A political theory tells individuals what they are morally required or permitted to do, or what they should refrain from doing, in their capacities as political actors. In our use, then, the term "political theories" denotes theories distinct from descriptive or causal theories, or explanations, of the rules and principles that political actors observe, or the values or institutions they embrace. Rawls's *A Theory of Justice* (1971, rev. 1999) is an example of a political theory in this normative sense. Alexis de Tocqueville's *Democracy in America* (1835–1840), in contrast, is an example of a (largely) descriptive political theory.[4] Most political philosophers are inclined to accept a political theory when it coheres with moral judgments that they find either intuitively compelling or acceptable for independent reasons; they reject the theory when it does not. If, for instance, a theory *T* entails that slavery is morally permissible, they reject *T*. This coherence, however, is not the only basis for assessing a political theory — or so we shall argue.

Let us say that political theories have both a *narrow* and a *broad* meaning. The narrow meaning is established by usage and exhausted by the set of normative judgments that the theory makes or entails. We shall say that a theory *justifies* those judgments. (Note that this terminological stipulation, adopted for stylistic convenience, allows that a false or otherwise unacceptable political theory can *justify* certain normative judgments.) The term "narrow" does not imply that this meaning is precise or unambiguous. But we assume that common usage sets empirically ascertainable constraints on a political theory's narrow meaning(s).

The broad meaning of a political theory consists in propositions stating the politically relevant causal consequences of political actors using the theory's narrow meaning, in conjunction with empirical claims, in practical deliberations about which political actors or public policies they ought to support. For example, suppose voters support candidate *C* because they endorse a theory of social justice that requires equal resources. Given their empirical beliefs, the voters perceive the theory to require a redistributive tax that *C* alone supports.

2. John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, rev. 1999), 175.

3. Ibid.

4. See Rawls (1999) and Alexis de Tocqueville, *Democracy in America*, trans. George Lawrence, ed. J. P. Mayer (New York: Doubleday, 1969).

The theory's broad meaning would then include statements of (a) this perceived requirement; (b) the causal impact of that perception on voters' willingness to support certain candidates or policies; (c) the causal impact of such willingness on the adoption of such public policies; and (d) the causal impact of adopting such public policies. An even fuller rendering of the broad meaning would include, inter alia, statements about legislators' and other political actors' views on the theory's additional implications for policymaking.

Let us also say that when political actors believe or publicly indicate that a theory provides normative reasons to adopt certain political decisions (presumably, in conjunction with certain empirical propositions), those actors *use* or *apply* that theory, provided they do so consistently with the theory's narrow meaning.[5] Thus, the voters we just imagined may well be using or applying their theory of justice. The broad meaning of a normative political theory, therefore, is determined by well-established propositions about the causal consequences of applying it. Our understanding of "using" or "applying" a theory is expansive, as it allows that the causal consequences integral to its broad meaning include political actors' publicly proclaiming it, perhaps to get re-elected, without believing that the theory necessitates or warrants political decisions that, uncoincidentally, benefit the proclaimer. *Pace* much political theorizing, we shall argue that a theory's broad meaning bears on its acceptability. Some examples and terminological stipulations will help us establish this thesis.

Suppose a political theory *T* entails that citizens have a right to quality college education. Assume, also, that this right entails that legislators are morally required to pass legislation granting high-school graduates admission to college upon their request. Finally, suppose that most voters either implicitly or explicitly endorse *T* and, predictably, believe that the right entailed by T requires colleges not to charge tuition fees — not even in the form of, say, state-financed vouchers. The voters believe this because they also believe that a quality college education requires direct subsidies. (We also allow that these beliefs are consistent with, even if not deducible from, *T*'s narrow meaning.) In this scenario, politicians would have an electoral incentive both to advocate subsidies to colleges and to publicly appeal to *T*, implicitly or explicitly, in order to justify those subsidies. Politicians might even have an electoral *dis*incentive to propose college vouchers, even if they would better secure the educational benefits *T* requires.

That the state ought to mandate colleges to admit high-school graduates upon request is part of *T*'s narrow meaning. The broad meaning of *T*, by contrast, includes the claim that legislation should provide for free college education without vouchers. Now suppose that *T*, together with well-established empirical propositions, actually justifies educational vouchers as uniquely appropriate to protect the right to quality college education entailed by *T*. While both this justification and the citizens' explicit or implicit reasoning rest on *T*'s narrow meaning, the citizens' variously articulated empirical beliefs lead them to reject, or at least ignore, a voucher system. As we shall soon see, the fact that both the voucher and direct subsidy proposals rest on *T*'s narrow meaning makes those proposals relevant to evaluating *T*. But we also allow *T*'s broad meaning — which includes the prediction that policymakers will subsidize college education without vouchers — to play a major role in that evaluation. We shall argue that this predicted application of *T*, as well as the resulting consequences, are part of the evidence for or against *T*.

A final stipulation before our main argument. We shall say that one *accepts* a political theory just in case one takes it to provide strong reasons for adopting certain public policies. An individual may accept a bad political theory — a theory that is unacceptable, in a sense of this term entailing that there are strong reasons for rejecting it. Since our main thesis concerns the conditions where a political theory is acceptable in this latter, objective-reason-based sense, we understand acceptability thinly enough to avoid theoretical commitments irrelevant to the argument. For example, we remain neutral about whether

5. This proviso is necessary to rule out critiques of a political theory based on predictable applications that flout any natural understanding of the theory's canonical formulations. See the last two paragraphs of Section II.

a theory's *truth* features essentially in the statement of reasons for accepting it, and about whether those must be *public* in a Rawlsian sense.[6] Our central thesis thus implies that predictable bad outcomes of political actors' accepting a political theory T are reasons against accepting (or, equivalently, reasons for rejecting) T. But we do not — and need not, given our limited goals — specify just how the strength of those reasons changes with the probability and (dis)value of the outcomes.

II. Content, Interpretation, and Acceptability of Political Theories

Why should one consider a political theory's broad meaning in deciding whether to accept it? Suppose that influential political actors are likely to invoke a political theory as a reason to adopt objectionable policies. The policies may be objectionable because of their outcomes. Or they may be inherently objectionable on deontological, symbolic, or other grounds. In conjunction with well-established empirical propositions, though, the theory by itself does not entail that those policies are either permissible or required. So how could appeals to that theory as a reason to adopt those policies count against the theory? Are we conflating the pragmatics of *invoking* a theory with the normative *content* of a theory? We will defend a negative answer, starting with an example that will provide initial intuitive support.

Suppose that, centuries from now, the guns in a political society — call it Futureton — have all been designed to guarantee that they can be put to a single use: deterring or defending oneself or others from would-be rights violators, such as rapists and murderers. With remarkable technological improvement, scientists in Futureton have created sensors for guns reliably to detect a situation's morally salient features. Those sensors disable the guns in cases of morally impermissible use. Guns in Futureton thus deter, and defend citizens from, would-be murderers and other rights violators, but do not facilitate gun-based crimes. That citizens of Futureton know about this happy improvement further undermines the criminals' threats of gun violence. Yet, the guns in Presentville are different. Like the guns in our real-life cities, these guns are used not only against criminals but also to facilitate murder, rape, robbery, and other rights violations. Gun use in Futureton is, in a natural moral sense, *better* than gun use in Presentville. After all, citizens of Futureton use guns only against rights violators, whereas citizens in Presentville use guns both against rights violators and to violate rights. We can extend this comparative judgment to other gun-related activities, such as buying and selling firearms. As these examples suggest, a proper assessment of guns and associated activities depends nontrivially on their likely uses, in a sense of "likely uses" that includes morally impermissible behavior.

Our central thesis extends to political theories the "likely-use" standard that we find so natural for assessing guns in this story. Guns and political theories that are less likely to be misused deserve a more positive evaluation. The Futureton guns are better than those in Presentville; after all, the former prevent more or worse evildoing than the latter. We propose a similar comparison of political theories: those that prevent (in a sense of "prevent" that also applies to the theoretical amendments to be suggested in section V) more or worse bad political decisions than rival theories do are in this respect better than those rivals.

We can spell out this idea by returning to the example of college vouchers. We imagined a political theory T that says citizens have a right to quality college education and most voters predictably take this right to require that colleges charge no tuition fees — not even fees payable with state-financed vouchers. We also imagined that T's broad meaning, which is determined by voters' reading of T and by their factual beliefs, induces policymakers to subsidize college education without vouchers. Let us now stipulate a further component of that broad meaning: most voters take T to give reasons to create an education agency that would determine which colleges qualify for subsidies. Now suppose such an agency would predictably waste citizens' money in red tape, award subsidies based on political party affiliations or other unacceptable criteria, and cater to special interests in ways that

6. See John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993).

produce social injustice, including severe shortages in quality college education. Assume, as well, that such bad outcomes would be much rarer under a voucher system. On these assumptions, we submit that *T* would be unacceptable.

Our critique of political theories runs as follows:

> (1) Suppose policymakers with non-negligible powers (*key political actors*, for short) will predictably use a political theory to adopt policies that cause some individuals to behave in ways that policymakers are morally required to discourage. Then, unless there are countervailing considerations, that theory is unacceptable.
>
> (2) Key political actors will predictably use political theory *T* to adopt policy *P*.
>
> (3) *P* causes some individuals to perform actions of type *A*.
>
> (4) Political actors ought to ban or discourage actions of type *A*.
>
> (5) No countervailing considerations override the (1)-based rejection of *T*.[7]
>
> Conclusion (from 1–5): *T* is unacceptable.

Call this type of argument the *Predictable-Use View*. Key political actors, as defined in (1), are best able to implement the principles and carry out the goals given in political theories. They are, in this regard, the primary addressees of those theories' normative messages. By assuming that they will predictably *use* a political theory, we mean to exclude misreadings of the theory's narrow meaning. This exclusion makes good sense. For it would be no objection to a political theory to say that key political actors will, even predictably, invoke it to bring about bad consequences, when that invocation is inconsistent with the theory's narrow meaning.

Here again the analogy with the gun example may help. We judge the likely use of guns in Futureton to be morally better than the corresponding use in Presentville, and we can extend the comparison to the guns themselves. In doing so it would be irrelevant to discuss knife crime rates in those two societies, unless we could also show that those rates depended on gun use. It would be a *non sequitur* to infer moral conclusions about the relative merits of those two types of gun use merely on the basis of knife crime rates. "Knife crime" differs in meaning from "gun crime." In the same way, policies that policymakers adopt because of a political theory *T* are irrelevant to *T*'s acceptability if the factual assumptions underlying those policies are irrelevant to — let alone contravene — the narrow meaning(s) of *T*.

III. Why We Are Not Conflating a Theory's Content with Its Application

Some may think that the Predictable-Use View conflates political actors' application of a theory with the theory's content. It is natural to think that one should accept or reject a theory on the basis of its content alone. Indeed, it is common to hear advocates from across the political spectrum dismiss as irrelevant allegations that their preferred political theories have been disproved by what supporters take to be misapplications of those theories. Many people who embrace principles they take to be egalitarian or socialist (say, the Marxian requirement that distributions be "from each according to their abilities, to each according to their needs"[8]) say that Stalinist regimes held mere ersatz forms of egalitarianism or socialism. And many people who accept principles they take as capitalistic, free-market, or libertarian, like those given in Robert Nozick's "entitlement" theory of "justice in holdings,"[9] say that crony-capitalist regimes violate those principles.

7. One countervailing consideration might be, for example, that enough individuals will predictably use *T* in ways that will bring about actions whose moral value outweighs the moral disvalue of actions of type *A*.

8. Karl Marx, "Critique of the Gotha Program," in Robert C. Tucker, ed., *The Marx-Engels Reader*, second edition (New York: Norton, 1978 [1875]), 531.

9. Robert Nozick, *Anarchy, State, and Utopia* (Oxford: Basil Blackwell, 1974), 151.

The general view underlying such moves involves what we may call

> *The Conflation Charge*. Political theories cannot possibly be undermined by violations of the rules and principles they propose.

We do concede that misapplications — understood as applications that canonical formulations of a normative political theory do not warrant — no more provide reasons against its truth, validity, or normative force than crime does against the criminal law's truth, validity, or normative force.[10] Our point is that a normative theory understood by competent speakers to require or authorize them to sanction or pursue independently objectionable outcomes is to that extent defective. Given that a normative political theory is inherently action-guiding, we ought to assess it by the actions or institutional arrangements that key political actors, who themselves are competent speakers of the theory's language, understand it to require or authorize.

We thus propose the following response to the Conflation Charge:

> 1. If key political actors will predictably use normative theory *T* to justify political decisions that will bring about independently objectionable outcomes, then *T* is, as it stands, defective.
>
> 2. Key political actors will predictably use *T* to justify political decisions that will bring about independently objectionable outcomes.
>
> Therefore,
>
> 3. *T* is, as it stands, defective.

It might be thought that premise 1 begs the question, since whether *T actually* authorizes or requires political actors to pursue measures bringing about independently objectionable outcomes turns entirely on *T*'s content. That is precisely the point of the Conflation Charge: misapplications of *T* distort that content. Hence they should not count against *T*, even if they are predictable.

This objection, however, overlooks the connection between normativity and action guidance. The guidance a theory provides can be conceptualized as the difference it makes in its addressees' practical reasoning. A political theory's guidance is thus revealed in its likely applications by competent speakers such as legislators, voters, and regulators. That is, the theory's guidance is revealed in its broad meaning, and *not* in its narrow meaning alone. The latter constrains the theory's broad meaning without making the two meanings coextensive. The broad meaning should count, then, in interpreting the theory's normative message, and ultimately count for or against the theory itself, given its normative nature. The next two sections develop this point.

IV. Conceptual Manipulation

In this section we discuss a major source of a political theory's predictable interpretations that count against it. Let us say that a political theory is *conceptually manipulable* when its narrow meaning allows many interpretations. Conceptual manipulation is common when two conditions are met: first, competent speakers disagree over how to interpret a term denoting a value *V* that the theory's narrow meaning deems centrally important; and, second, some key political actors' attempt to promote or respect *V* under some interpretations will be detrimental either to *V*, whether on the same or a different interpretation, or to other values at least as significant as *V*. We next show what is wrong with conceptually manipulable political theories, and how the Predictable-Use View explains their flaws.

Let us begin with Isaiah Berlin's classic discussion of liberty as a political ideal.[11] Berlin distinguishes between negative and positive

10. Readers uncomfortable with the idea that the criminal *law* can be true or false may substitute *descriptions* or *descriptive theories* of the criminal law, such as those found in casebooks.

11. Isaiah Berlin, "Two Concepts of Liberty," in Isaiah Berlin, *Liberty: Incorporating Four Essays on Liberty*, ed. Henry Hardy (Oxford: Oxford University Press, 2002 [1958]).

conceptions of liberty. Negative conceptions hold that an individual is free just in case no one else prevents her from doing what she otherwise could do. Berlin sees commitment to negative liberty in classical liberals like John Locke and Adam Smith, who believed that "social harmony and progress were compatible with reserving a large area for private life over which neither the state nor any other authority must be allowed to trespass."[12] But Berlin also notes that the term "liberty" (or, equivalently, "freedom") takes on a different meaning in later political theory, and he calls this meaning "positive." An individual has positive liberty just in case she is her own master and not under the control of external (human or non-human) forces. Berlin argues that the positive conception of liberty naturally views the self as divided. A person's autonomous, noble, and rational self has desires and ends that are truly hers and, therefore, not those of her enslaved, base, and irrational self. This latter self's desires do not express that person's *true* nature. Her higher, genuine self, in fact, desires liberation from them. This positive conception of liberty allows one to say, for instance, that a drug addict lacks liberty insofar as addiction enslaves him, even if no one else interferes with his choices (i.e., the choices of his lower self). The addict has negative liberty but lacks the self-control that positive liberty entails.

Berlin observes that authoritarian regimes have appealed to positive conceptions of liberty to justify appalling intrusions into individuals' lives. These regimes' rhetoric has been used to justify police states aimed at "liberating" citizens from their "base inclinations." And, Berlin adds, once the original negative meaning of "liberty," which had long prevented speakers from saying that a police state promotes or respects liberty, mutated into the positive sense, calls for massive governmental coercion were no longer perceived as inherently hostile to liberty. On the contrary, they could now be made *on behalf* of liberty. Here is Berlin:

> The rationalist argument [for the unlimited political power demanded by self-proclaimed liberators of individuals' nobler selves], with its assumption of the single true solution, has led by steps which, if not logically valid, are historically and psychologically intelligible, from an ethical doctrine of individual responsibility and individual self-perfection to an authoritarian State obedient to the directives of an élite of Platonic guardians.[13]

Now, suppose we can predict that a theory of positive liberty will frequently be used to support illiberal policies. If our argument so far is correct, and if illiberal policies are themselves morally objectionable, then this prediction would undermine that theory of positive liberty. We could test this prediction by ordinary empirical methods, such as analyses of the correlation between, on the one hand, governments' public advocacy of the ideal of positive liberty or their denunciation of negative liberty as spurious, and, on the other, their repressive policies. Alternatively, we can ground that prediction by reflecting on the sorts of argumentative moves a commitment to positive liberty invites, as we do next.

Berlin's discussion does not show that the positive conception of liberty fails on its own terms. Indeed, the positive liberty theorist (henceforth, "positive liberal") might say that the allegedly illiberal policies Berlin denounces, far from rendering her own position objectionably self-defeating, in fact realize the value of true, positive liberty. Moreover, Berlin gives us no reason to believe that the negative conception of liberty is entirely immune to authoritarian manipulation. A government might curtail an individual's negative liberty, for example, and defend that curtailment by saying that it removes interferences with the individual's *genuine* desires, which flow from her "higher" self. *Both* conceptions of liberty are manipulable, then, as long as they distinguish between higher and lower selves.

12. Berlin, "Two Concepts of Liberty," 173.

13. Berlin, "Two Concepts of Liberty," 198.

Now Berlin says that the question for positive liberals is, "What, or who, is the source of control or interference that can determine someone to do, or be, this rather than that?"[14] This question concerns the identity of the bearer of positive liberty. The question leaves open the possibility that this bearer of positive liberty is not the ordinary, directly observable, individual, but someone with a deeper self that political functionaries should discover and developmentally assist. The question therefore allows determination of the appropriate scope of coercive political action to pivot on empirically untestable claims about the nature of persons and their desires. Even equally rational, well-informed people, or "epistemic peers," might disagree radically — perhaps because they select, order, or weight their reasons differently before arriving at conclusions.[15] If so, then inasmuch as the epistemic peers disagree, illiberal governments can easily invoke expert support for claiming that they are serving citizens' empirically untestable desires. This will be the case even if others find their conclusions about proper government action mistaken, and even if, in fact, these governments are knowingly disserving citizens. The positive conception of liberty thus enables illiberal governments to manipulate appeals to a higher self on behalf of liberty. Put differently — in a world where governments are empowered to promote positive liberty, and rational and well-informed citizens and policymakers have different political preferences, governments do not need to overstep the narrow meaning of canonical formulations of ideals of positive liberty in order to use coercion to further their own ends. This holds regardless of whether any attractive political morality would sanction those ends.

14. Ibid.

15. See Thomas Kelly, "The Epistemic Significance of Disagreement" (2005), in T. S. Gendler and J. Hawthorne (eds.), *Oxford Studies in Epistemology*, Vol. 1, Oxford: Oxford University Press, 167–196, and Robert Mark Simpson, "Epistemic Peerhood and the Epistemology of Disagreement" (2013), *Philosophical Studies* 164(2), 561–577. On cross-cultural disagreement, see Gregory Robson, "Magistrates, Mobs, and Moral Disagreement: Countering the Actual Disagreement Challenge to Moral Realism" (2021), *Canadian Journal of Philosophy* 51(6), 416–435.

For example, imagine that citizens have different interpretations of their own higher selves as well as those of others. Some view those selves as endorsing traditional family values, others a life of religious seclusion, and still others entrepreneurship. Some view their higher selves as valuing a medical career more than an artistic one, whereas others hold opposite views. Let us further assume that everyone regards positive liberty as highly valuable, and is aware of the stark differences between the desires of the higher selves respectively embraced by the citizens. Governments could therefore select, on a case-by-case basis, among those interpretations of the higher self, or among ways of trading them off against each other, that best publicly justify their policies. Such an erratic application of a theory of positive liberty may cumulatively reduce citizens' ability to further their life plans, and so might actually thwart their positive liberty, at least in the eyes of those favoring alternative tradeoffs. Yet governments could do all of these things without facing the charge that any one of their policies failed to prioritize positive liberty. After all, government has traded it off, when necessary, only against itself.

The relevance of likely applications of a theory to its acceptability, defended previously, entails that when governments predictably use the theory of positive liberty this way, that theory, *as it stands*, is unacceptable, absent countervailing considerations. In the next section, we discuss the nature of the amendments needed to rescue theories whose predictable uses render them defective as sources of moral guidance in political decision-making.

V. Theory Change Under the Predictable-Use View

We have argued that the predictable bad consequences of likely applications of a political theory count against it. What methodological advice follows? We propose a

> *Suggestion for Theory Construction*. Theorists should replace any political theory *T*, whose broad meaning includes an objectionable normative message *N*, with another theory,

> *T**, that retains the plausible components of *T*'s broad meaning while ruling out *N*.

*T** might include, for instance, a proviso to ensure that *T*'s applications will not be affected by perverse incentives. Suppose, for example, that we have strong evidence to believe that authorizing governments to promote positive liberty will, as we imagined in section IV, give rise to despotic rule. *T** might accordingly forbid such authorizations.

Here is a more specific example of such a proviso. Suppose that likely applications of Rawls's difference principle will be counterproductive — pretend that the scholarly literature has given good evidence for that prediction. We could address the issue by adding the following proviso to standard formulations of that principle:

> This principle may not be used to justify redistributive taxation that would significantly increase unemployment due to moral hazard, reduce investment in industries whose workers and consumers are among the least well-off in terms of primary goods, or, more generally, bring about effects that offset the benefits of redistributive taxation for the least well-off.[16]

We can suitably amend the difference principle to create the above version or others, as determined by well-established hypotheses about the factors leading to counterproductive applications of the principle. Such provisos might well make political actors less likely to adopt bad policies.[17]

All else equal, we believe a qualified version of a political theory is warranted when the source of the objections to its broad meaning

16. On the welfare losses that investment-unfriendly taxes inflict on the poor, see David Schmidtz, *Elements of Justice* (Cambridge: Cambridge University Press, 2006), 140–149.

17. "Less likely" because nothing here is failsafe. Someone might have false empirical beliefs about the qualified difference principle that support predictable misuse—even if, as suggested below, adding provisos is a self-correcting process. Thanks to an anonymous reviewer for pressing us to qualify the accompanying text.

is not, as we imagined with the difference principle, a causal mistake, but is instead what we might call *interpretive confusion*. An interpretive confusion is an interpretation, or sequence of interpretations, of key concepts used in the theory's canonical formulations that, despite being consistent with the theory's narrow meaning(s), still licenses independently objectionable policies. We illustrated such a policy with Berlin's warnings about appeals to positive freedom in order to justify totalitarian regimes. Our Suggestion for Theory Construction would require political theories centered on the value of positive freedom to explicitly embrace values such as political pluralism, free speech, due process of law, the universal franchise, and other safeguards against objectionable appeals made to the positive freedom ideal. The Suggestion would also require suitable, non-manipulable formulations of those ideals and institutions.

Here, some readers may pause and observe that a political theory does not, on its own, prescribe independently objectionable behavior. To do so it must connect with factual propositions. The falsity (or unacceptability) of the policy or institutional recommendations derivable from that conjunction would then falsify *either* the political theory *or* the factual premises. But then, the objection may conclude, our central thesis would founder, for it would arbitrarily highlight the theory's role in generating those recommendations, even if the factual premises were false.

A related problem is that the Predictable-Use View seems to generate reasons against *any* political theory, including true theories. Any factually mistaken agent who applies *T*, a true theory, and as a result brings about independently bad outcomes, would on the Predictable-Use View thereby provide *some* reason to reject *T*. Put differently, the Predictable-Use View seems to overgenerate reasons to reject true theories. To be sure, a theory's truth is consistent with there being reasons to reject it, because reasons for accepting it may outweigh them. But it would seem that such cases render the Predictable-Use View too unreliable.

We think that worries like this can be dealt with by imagining situations in which a true normative theory makes a decisive practical difference in an agent's disastrous choice: where the agent would not have made that choice had they not come to accept that theory, and they then make the choice due to an empirical mistake. Given the severity of the outcomes, and assuming that the agent uses the theory (i.e., their practical reasoning conforms to the theory's narrow meaning), we might be tempted to say that the Predictable-Use View should take the broad meaning expressed by the agent's choice as especially troublesome for the theory, even though it is true by assumption. This in turn would seem to undermine the Predictable-Use View: it would generate reasons (but perhaps nonconclusive ones) for rejecting true theories.

The following sample reasoning illustrates the problem:

> 1. Agents have an overriding moral duty to maximize expected aggregate happiness.
>
> 2. Jessica would maximize expected aggregate happiness just in case she pressed this red button right here, which she had never seen before, behind the fridge in her Queens apartment.
>
> Therefore,
>
> 3. Jessica has an overriding moral duty to press this red button right here.

Suppose that 1 is true (in an objective, belief-independent sense of "true"[18]) and 2 is false. Sadly, pressing the button would trigger a military attack and spark a world war with immense avoidable suffering. Jessica believes 2 and has the button within reach. She also believes, correctly — or at least would come to believe on reflection — that 3 follows from 1 and 2. And she is motivated to do what she takes to be her moral duty. For whatever reason, though, Jessica has no views about 1, and so does not believe that she has more reason to press the button than not to. Still, she is equally open to reasons for or against 1.[19] *Under these circumstances*, it would surely be wrong for us to persuade Jessica of 1's truth — doing so would lead her to press the button! Our assumption that 1 is true clearly gives us a reason to persuade her of 1's truth *and* the falsity of 2. Indeed, in these circumstances, we would have no moral reason to persuade Jessica of the truth of 1 *without also persuading her of the falsity of 2*. The Predictable-Use View coheres with this judgment because it entails that the guidance from a political theory to ill-informed agents may count against the theory, unless that guidance is qualified as noted in the Suggestion for Theory Construction.

Another feature of the Jessica case, however, seems to support an earlier critique. We assumed that 1 is true, and yet the Predictable-Use View appears to take 1's guidance for Jessica (given her beliefs and motives) as a reason to reject 1. 1 may still be true, all things considered, yet it seems problematic for the Predictable-Use View to imply that cases such as Jessica's generate a reason to reject a principle — here, the principle of utility in 1 — that is true by hypothesis. More generally, the Predictable-Use View seems to overgenerate reasons to reject any true normative theory. For we could produce those reasons by giving suitable false beliefs to those willing and able to use the theory.

Recall, though, that the Predictable-Use View claims that the bad consequences of *policies,* predictably adopted by *key political actors* on behalf of a political theory, constitute a reason to reject that theory (see Section II). For all we know, Jessica is not a key political actor — she

18. For discussion of how a duty might be objective, see Judith Jarvis Thomson, *Normativity* (Chicago: Open Court, 2008), 188–191. The objectivity of the duty proclaimed by 1 gives us, as we note below, a reason to persuade Jessica of 1's truth (and 2's falsity).

19. It does not matter here why she has no views on 1. She may be uncertain of its truth, be a moral skeptic open to non-skeptical arguments, or have some reason for suspending judgment on 1 and so on whether she ought to press the button. We need only assume that she is responsive to reasons for revising her current suspension of judgment on 1. Under the specified circumstances, this assumption ensures that, as noted before, coming to believe 1 makes a decisive difference in Jessica's practical reasoning.

does not face the choice in her capacity as a voter, legislator, or other important policymaker. Moreover, since she faces a choice between a particular action and its omission — i.e., pressing the button or not doing so — her behavior could not have been predicted under the description "adopting policy *P*." Even in a variant of the case where she did face that choice in her capacity as a key political actor, she would not be facing a choice between public policies; pressing the button and not doing so are not public policies. Under these circumstances, her using the principle of utility — were we able to convince her of its truth — as a reason to press the red button, and thereby bring about disastrous consequences, lies beyond a predictable-use critique of that principle. Put differently, propositions describing Jessica's choice and its consequences — even in the political variant of the case we just imagined — are not part of the principle of utility's broad meaning. This follows from our definition of a political theory's broad meaning (section I) in terms of propositions stating the politically relevant causal consequences of key political actors' using a theory in their practical deliberations about the political actors or public policies they ought to support. It would be a mistake, then, to take the Jessica case as suggesting that the Predictable-Use View overgenerates reasons against political theories, even if those theories are true.

It might be objected that we are confining the scope of predictable-use critiques ad hoc, in order to prevent the Predictable-Use View from overgenerating reasons against political theories. We would then be arbitrarily confining predictable-use critiques to cases in which key political actors adopt bad public policies. Why wouldn't *particular* misuses of moral principles by *non-political* agents, such as Jessica, count against those principles, especially if those misuses were predictable? Suppose there is no nonarbitrary way to distinguish the Jessica case from the political cases to which the Predictable-Use View applies. Then the Predictable-Use View would seem committed to overgenerating reasons against all kinds of moral principles — not just the principles of political morality proposed by political theories.

We contend in reply that predictable-use critiques are especially fitting to political behavior. Key political actors have an incentive publicly to claim that the values or rights in attractive political theories will be best promoted or protected by adopting certain public policies — those that most citizens perceive as *directly* responsive to the values or rights. This discursive behavior is the instrumentally rational response to most citizens' disincentive to invest much time or money in studying the complex, and often counterintuitive, causal connections between public policies and the outcomes required by their chosen principles of political morality.[20] To be sure, the perceived directedness of the policies adopted on behalf of values or rights may result from more or less implicit commitments to symbolic or deontological conceptions of practical rationality. These conceptions may place less weight on the causal consequences of policies, especially if they are less direct.[21] But the tiny probability that an ordinary citizen will find ways to affect public policies makes it instrumentally rational for most citizens not to investigate the causal consequences of alternative public policies. This is true even if those citizens believe that the consequences are relevant, yet not necessarily decisive, to their supporting or resisting a particular policy. We should expect public political discourse, as a result, to *sound* more symbolic, deontological, or otherwise consequence-insensitive than warranted by the weight that the actual causal consequences of public policies would have in citizens' assessments of those policies (if they had an incentive to get the information). We take no sides here on whether and when such apparent consequence-insensitivity is warranted. But we do assume

20. The seminal formulation of the "rational ignorance" hypothesis is Anthony Downs, *An Economic Theory of Democracy* (New York: Harper, 1957). For fuller discussion of our remarks here, see Guido Pincione and Fernando R. Tesón, *Rational Choice and Democratic Deliberation: A Theory of Discourse Failure* (New York: Cambridge University Press, 2006).

21. Notice that this says merely *less* weight. As Rawls emphasized, "All ethical doctrines worth our attention take consequences into account in judging rightness. One which did not would simply be irrational, crazy." See Rawls (1999, 26).

that *some* attention to consequences is a plausible requirement on correctly applying principles of political morality. We assume, for instance, that any defensible rights-based political morality should be sensitive to morally relevant consequences of alternative ways to protect rights. It should also be sensitive to consequences catastrophic enough to warrant minor infringements of persons' rights, especially if due compensation is paid to the victims. And we also assume that most ordinary citizens are consequentialists in this weak sense of the term.[22] As long as political theories govern citizen behavior of that sort, as well as policymakers sensitive to the citizens' moral and causal views, our proposed scope of predictable-use critiques of political theories is appropriate.

Now to say that such critiques are appropriate in the field of political theory is not to deny that they are appropriate in other fields as well. We take no stance on this question, other than to note that as long as persons occupy roles in nonpolitical organizations — say, as CEOs, journalists, or doctors — the claim that broad meanings matter in selecting normative theories for those roles would be true, given a suitable redefinition of "broad meaning" to cover the behavior of the role occupants. We accordingly expect good theories of nonpolitical role obligations to include provisos ruling out certain policies. Thus, like our proposed provisos on political theories, these provisos should be sensitive to the epistemic and motivational constraints under which those occupying roles in hospitals, healthcare regulatory agencies, and other organizations adopt policies regarding controversial policies such as voluntary euthanasia. Normative theories of such roles could be vulnerable to critiques based on "misapplications" in much the same way as political theories would. While theories for or against *legalizing* voluntary euthanasia are political theories in our sense of the term — and, as such, are potential targets for predictable-use critiques — similar critiques of the theories of professional roles we are now imagining require redefining "broad meaning" in order to make those critiques apply to actions performed, or policies adopted, by role occupants.

The Jessica case does not involve these roles. Jessica is an unspecified agent. None of the provisos constraining the range of policies the principle of utility allows legislators, voters, and other key political actors to adopt or support are germane to the one-shot choice she faces. It is true that her coming to believe the principle of utility makes a big difference under the stipulated circumstances. That principle becomes the only source of guidance for her in a situation where, as we imagined, she has no overriding reason to press the button or not. But as we pointed out, that guidance is not part of the principle of utility's broad meaning, since she is not a key political actor — her choice whether to press the button is not one she makes in that capacity. The Predictable-Use View would not take her pressing the button as a reason to reject the principle of utility. It does not even take that choice as a *pro tanto* reason to reject the principle. And that is as it should be. After all, the bad consequences of using a normative principle should not decisively undermine that principle in situations where, unlike much political decision-making, the agent's behavior manifests no systematic tendency to bring about bad consequences. The bad consequences do not show, in particular, that the principle as it stands is less acceptable than some appropriately qualified version. For the uniqueness of Jessica's situation prevents us from saying that *agents in a specifiable type of situation*, which she happens to instantiate, tend to err in systematic ways. As a result, we have no basis for formulating a proviso on the principle of utility that addresses *those* situations. In conclusion, the Jessica case does not show that the Predictable-Use View overgenerates reasons, even *pro tanto* ones, to reject true theories.

Our restricting the Predictable-Use View to political theories need not presuppose views about political decision-making and its accompanying rhetoric that some readers may find overly cynical. Our argument is compatible with fairly optimistic views about the nature of

22. See Guido Pincione and Fernando R. Tesón, "Self-Defeating Symbolism in Politics," *The Journal of Philosophy* 98 (12), 2001: 636–652, and Guido Pincione and Fernando R. Tesón, "Rational Ignorance and Political Morality," *Philosophy and Phenomenological Research* 72 (1), 2006: 71–96.

political action and discourse, at least in stable liberal democracies. The view's practical relevance as a source of critiques of political theories presupposes only that key political actors have incentives to adopt bad policies on behalf of attractive values and principles, and consistently with their narrow meanings. It does not rule out that, more often than not, they overcome those incentives. As long as the broad meanings of the principles or values in political theories involve adopting bad public policies, the Predictable-Use View generates sound objections to those theories, and its associated Suggestion for Theory Construction shows a way to theoretical progress.

This remark on theoretical progress is worth spelling out, for we do not merely claim that the Predictable-Use View makes room for a largely overlooked type of reasons to reject political theories. We also contend that the Suggestion for Theory Construction leads to two types of improvements in political theories, including those qualified by its recommended provisos. On the one hand it creates a self-correcting process. As illustrated by the proviso on the difference principle we imagined earlier, one rationale for the provisos is the prediction that policymakers will misapply political theories as a result of factual mistakes. We saw next that another major rationale is interpretive confusion, which we illustrated with grossly illiberal policies proposed in the name of positive freedom. In both examples the provisos aimed to forestall objectionable applications of political theories. The provisos turn on hypotheses about the probabilities of key political actors' adopting certain policies, and of those policies' bringing about bad outcomes. These hypotheses may well be false — in particular, it may be false that a proviso intended to prevent a certain outcome will be read, perhaps by causal mistake, as authorizing or even requiring policies bringing about that outcome or something even worse. Such falsities would not undermine the Predictable-Use View. Rather, they would undermine particular attempts to follow the Suggestion for Theory Construction. The Suggestion for Theory Construction itself would tell us to replace the defective provisos with more reliable or acceptable ones, consistently with the narrow meaning of the political theory they qualify.

On the other hand, the provisos often have great heuristic value: they deepen our understanding of the theories they qualify, even in ways that speak against those provisos. What we have in mind here differs from the provisos' role in preventing misapplications of political theories. Imagine that a political theory (T) asserts or entails that legislators ought to adopt policies enhancing the dignity of the least well-off. For the sake of illustration, assume two more things: (i) predictably, legislatures will use T or its principles as justification for raising the minimum wage above market wages; and (ii) also predictably, the raises will increase unemployment among unskilled workers in industries where market wages are below the new minimum. The unemployment increases are, let us also assume, independently bad. As such, they can plausibly be deployed in a predictable-use critique of T as well as figure in support of another theory, T^*, that consists in the conjunction of T with a proviso excluding minimum wages from the set T's permissible applications. Finally, we can now assume that T^*, and in particular its proviso on T, overlooks the most compelling reason for the minimum wage — that it prevents the exploitation of workers. This objection might be the basis for a second reformulation of T, T^{**}, which fleshes out the idea of enhancing dignity for the least well-off in a way that entails that fighting exploitation is morally more important than fighting unemployment. T^{**} thus does not ban minimum wages. Now some plausible predictable-use objections to T^{**} itself might be imagined. For example, some critics might adduce that unemployment gives rise to more serious forms of exploitation than underpaying workers.[23]

This series of moves and countermoves might last for a long time, and we defend no particular argument in the series. But at some point a proviso, by barring policies that seem to meet independent moral requirements, can reveal the need to revise the theory it qualified.

23. For a discussion of possible reconstructions of philosophical arguments for the minimum wage, see Pincione and Tesón 2006.

Ideally the revisions will leave us with a theory capable of explaining the elimination of, or changes in, that proviso. Thus, *T**, and especially its proviso against minimum wages, was heuristically instrumental in developing *T***. *T***, by its reference to exploitation and the absence of a proviso against minimum wages, was in turn heuristically instrumental in refining our understanding of the relationships between unemployment and exploitation, perhaps even in ways leading to objections to *T*** itself. We cannot overstate the importance of the provisos' heuristic role, including their elimination. For we can now better understand the sense in which the Predictable-Use View is self-correcting. While its verdicts are not failsafe, the provisos given by the Suggestion for Theory Construction help us to identify false or otherwise defective theories — including those resulting from prior applications of that suggestion, as well as the kinds of considerations that improved theories will have to accommodate.

VI. Concluding Remarks

The Predictable-Use View takes predictable applications of a political theory as evidence against it when those applications are independently objectionable. That evidence may warrant us in adding public policy qualifications to the theory's original formulation. We emphasized that, more often than not, such qualifications do not merely rule out certain public policies. They also change the *original normative message of the theory*, or the difference it makes to the practical reasoning of political actors given their causal and moral assumptions. Those assumptions include beliefs about the existence and weight of factors other than those that figure in the theory. To that extent, such policy qualifications are theoretical improvements: they either improve the guidance that the original, unqualified version of the theory gave to key political actors, or they play a heuristic role in reformulating the theory to accommodate the weight of alternative principles or values.[24]

24. We thank Horacio Spector for helping us to clarify and develop key ideas in this article. We also thank an anonymous referee for detailed and insightful criticisms on earlier drafts.