The Logic of Idealization in Political Theory

Short Title: Logic of Idealization in Political Theory

Author: Jonathan Leader Maynard

jonathan.leader_maynard@kcl.ac.uk

Department of Political Economy, King’s College London, 30 Aldwych, London, WC2B 4BG, United Kingdom

Key Words: Idealization, ideal theory, nonideal theory, idealism, models, counterfactuals, logic of inference, methodology, robustness, normative theory

Acknowledgements

My thanks to Adrian Blau, Keith Dowding, Steven Klein, Laura Valentini, the AJPS editors, three anonymous reviewers, and participants in the ‘Methods of Normative Political Theory’ section of the ECPR Annual Conference 2023, for invaluable feedback on earlier drafts of this paper. My thanks also to David Wiens for sharing a draft of his book manuscript, *The Best and the Rest: Thinking with Ideals in a Non-Ideal World*, while I was revising this paper.
The Logic of Idealization in Political Theory

Abstract

The role of ideals and idealizations is amongst the most vigorously debated methodological questions in political theory. Yet the debate seems at an impasse. This paper argues that this reflects a fundamental ambiguity over idealization’s intended inferential logic: the precise way in which idealizations might yield normative knowledge. I identify two tacit understandings of idealization – a dominant ‘telic’ understanding, and a less overt ‘heuristic’ understanding – which, though importantly different, are rarely distinguished. I argue that delineating these understandings, and shifting from telic to heuristic idealization, recasts various unresolved methodological problems for political theorists, while productively connecting their discussions to work on idealization in political science and the practice and philosophy of science more broadly. I then provide a systematic account of how idealization might be used heuristically in normative reasoning and explicate the advantages of such approach for promoting rigorous, relevant, and inclusive methodologies in political theory.

Word Count: 9949
The Logic of Idealization in Political Theory

The role of ideals and idealizations (and, conversely, of empirical facts and real contexts) represents one of the most vigorously debated methodological questions in political theory. The distinction between ‘ideal’ and ‘nonideal’ theory, and the contention that theorizing ideal justice is a precondition to working out what to do in real (nonideal) circumstances, was one of John Rawls’ central contributions to political theory methodology.¹ Many political theorists continue to present the theorization of ‘ideal’ or ‘perfect’ justice as a central task of the discipline (Swift 2008; Stemplowska 2008; Stemplowska and Swift 2012; Estlund 2014, 2020; Bertea 2023; Sirsch 2021). Yet a focus on ideal conditions, the neglect of nonideal realities, and the very distinction between ideal and nonideal theory, have all been heavily criticized by opponents of ‘Rawlsian’, ‘utopian’, ‘unpolitical’ or otherwise ‘ideal’ political theory (Mills 2005; Sen 2010; Wiens 2015, 2024; J.T. Levy 2016; Gaus 2016; Hall 2016; Rosenberg 2016; Sleat 2018).

Despite an outpouring of commentary, this debate seems at something of an impasse: with defenders of idealization practicing political theory much as they always have, while critics castigate such inquiry as wrongheaded, useless or worse. Besides exacerbating disciplinary fragmentation, this has substantive implications, because questions of how ‘idealistic’ or ‘realistic’ our politics should be, and the relevance of behavior in idealized circumstances for real-world decisions, have downstream consequences for one’s normative conclusions. Such questions are not limited to debates surrounding ‘Rawlsian’ political philosophy. They overlap with the recent revival of ‘realist’ political theory (e.g. Hall 2016; Favara 2023),² and, as

¹ These claims had earlier precedents, see: Gaus (2016, 2-4); Estlund (2020, xi).
² The debates over idealization and political realism are distinct, however – see: Sleat (2016).
Johnson (2014) observes, theorists working in the more ‘critical’ or ‘continental’ tradition also employ idealized models, such as Foucault’s (1977) ‘panopticon’ or Habermas’ (1981/1987) ‘unlimited communication community.’ Nor are these questions purely scholarly. “Many people even outside academic debates,” Estlund (2020, 4) observes, “have strong views and deep questions about what it would mean to put idealism and realism about politics in their proper place.”

In this article, I propose a way to rethink this debate. I argue that both defenders and critics of idealization have often remained vague on the inferential logics through which idealization might help justify normative claims. In particular, scholars have not delineated two distinct presentations of idealization’s role. The first construes idealization as a means of modelling the ideal state of some value – such as justice – that provides the goal, target, or benchmark from which other (nonideal) normative conclusions are derived. The inferential logic is that any normative judgment in the nonideal real world presupposes some such ultimate normative ideal in which the value is perfectly or fully realized (Simmons 2010). This telic understanding, as I term it, is the dominant current portrayal amongst many defenders, and almost all critics, of idealization in political theory.3 Yet there is a second, heuristic, way of understanding idealization: in which idealized models are tools for analyzing and interrogating normative standards, values, or principles,4 rather than deriving them from an ideal end-state. A few writers aver to this more heuristic understanding of idealization. But it is a much less common explicit portrayal, and rarely figures in extant methodological debates.

3 Rawls’ view, however, is ambiguous – see footnote 14 below.
4 ‘Standards’, ‘values’ and ‘principles’ are not identical, but for simplicity I will use ‘principle’ as a placeholder term for all three concepts.
Because these two understandings of idealization are rarely distinguished, they represent rough ways of thinking and talking rather than a well-specified distinction commonly understood by idealization’s defenders and critics. I argue that properly delineating them, and shifting from the telic to the heuristic account, recasts a range of unresolved problems surrounding the place of ‘ideals’ in political theory. I will not definitively refute telic idealization. But I suggest that in generally fixating on it, defenders and critics have neglected a heuristic understanding that offers clearer reasons to see idealization as crucial in normative thinking. That understanding avoids most of the problems critics have identified with excessively ‘idealistic’ or ‘utopian’ political theories, while rendering idealization in political theory more consistent with prevailing understandings of idealization in the social and natural sciences and philosophy of science. It also appropriately reframes most remaining disputes about the place of ideals in politics as substantive ethical disagreements rather than matters of technical methodology.

The claim that idealization in political theory is heuristically useful is not unprecedented. List and Valentini (2016, 544-6), Ismael (2016), and Nefdt (2021) all suggest that idealization in political theory is comparable to idealization in science, where idealized models represent useful tools. J.T. Levy (2016) powerfully critiques the very notion of ideal theory yet accepts that idealizations can remain instrumentally useful. Favara (2023) endorses heuristic idealization, in my terms, but presents it as distinctive to ‘realist’ political theory, while seeing non-realists as committed to what I call telic idealization. Wiens accepts that idealized models in political theory can be useful for conceptual purposes but contends that they are “useless for thinking about normative matters” (Wiens 2024, 4). I build on the contributions of all these

5 Like most participants in this debate, I am interested in action-guiding normative conclusions (see: Gaus 2016, 11-18; Barrett 2023, 33). I set aside the argument that idealization may provide theoretical truths even if these lack action-guiding implications (Cohen 2008; Estlund 2020).
6 There is also a growing literature on thought experiments, case comparisons and models in political theory (Johnson 2014; Brownlee and Stemplowska 2017; Slavny, Spiekermann, Lawford-Smith, and
scholars. None of them, however, draws the distinction between telic and heuristic idealization I make in this paper. Nor do they provide a systematic account of how idealization might be used heuristically in normative reasoning across different methodologies or explicate the implications this has for broader debates over ideal theory and idealization. That is what I contribute here.

I proceed in four main sections. Section one briefly sketches the present ambiguity over the meaning of ‘ideal theory’ and ‘idealization’ in political theory and the inferential logics they involve. Section two then lays out the dominant telic understanding of idealization and the theoretical difficulties it has faced. These two sections are intentionally somewhat summative but provide important foundations for the more novel arguments developed in the rest of the paper. Section three then lays out the alternative heuristic understanding of idealization I defend, explicating a key inferential logic – counterfactual coherence – that underpins its relevance for normative reasoning and justification. Section four then outlines the advantages and implications of shifting to a heuristic understanding of idealization for broader debates over ideals and idealization in political theory.

1. The Ambiguity of Idealization

The central methodological question surrounding idealization in political theory is effectively: *what relevance do depictions of idealized scenarios have for normative (political) judgments?* As originally construed by Rawls (1971, 8-9 & 245-6), ‘ideal’ and ‘nonideal’ theory were crucial but different tasks. Ideal theory identifies our underlying normative principles by imagining scenarios of full compliance (where everyone will do the right thing once identified)
and favorable conditions for justice (where meeting everyone’s basic needs is feasible). Nonideal theory then analyses how the principles identified through ideal theory should be pursued in nonideal conditions with obstacles for ideal justice, including people’s noncompliance with what justice requires of them. For Rawlsians, this construal reflects an intuitive claim: that while people failing to do the right thing alters the normative context of action, normative requirements do not cease being requirements just because people refuse to follow them (Estlund 2014; 2020).

Rawls’ construal faces an immediate problem, however: that this binary distinction between ideal and nonideal theory seems untenable. As J.T. Levy (2016) argues, all normative theorizing must assume some ‘moral friction’ – some problematic states of affairs – to even get going. Hamlin and Stemplowska (2012, 50) similarly observe that if ideal theories strictly assume ‘full compliance’ with ideal principles, there would be no problems of institutional design for ideal theory to work on. As such, Levy concludes, there is “no such thing as ideal theory.” All normative theories hold some features of reality fixed, normatively troublesome, and not appropriately subject to idealized removal, while treating other features of reality as mutable and something our normative theories might require us to change.

The counterpart to this observation, however, is that all normative theories idealize to some degree: they must model scenarios that are different from (and typically somewhat preferable to) our present reality (J.T. Levy 2016, 314-15). Considering counterfactuals is an inescapable part of justifying any action-orientated claim – one must consider how things could be different (see also: Goodin 1995, 43-4; Estlund 2020, 5). Normative theories must therefore work with
some models, however tacit, of at least somewhat idealized worlds.\textsuperscript{7} We might limit such models to very proximate plausible scenarios, draw on our best empirical knowledge, and also analyze historical cases or real-world experiments (Schmidt 2011, 788-9). But no existing empirical case or experiment provides an exact analogue to possible future cases we normatively contemplate. Categorical rejections of idealization are therefore misplaced: employment of idealized models is a ubiquitous feature of human inquiry that cannot be avoided entirely (Stuart 2022; Mäki 2020; Appiah 2017, 104-11). Different theories simply idealize in different ways and to different degrees (Hamlin and Stemplowska 2012; Valentini 2009, 333; J.T. Levy 2016, 314).

I emphasize that this really is ‘idealization’ and not merely ‘abstraction’. Some political theorists attach considerable weight to this distinction: accepting that all theories abstract (since they simplify reality) but not that all theories idealize. Yet the distinction is knotty, because the common claim (e.g. O’Neill 1987; Hope 2016, 376-7; List and Valentini 2016, 545) that abstractions merely ‘leave out’ aspects of reality whereas idealizations make ‘false assumptions’ is unsustainable, as generally recognized in philosophy of science. For a start, it is misleading to tie idealization to the language of ‘false assumptions,’ because idealizations are not erroneous propositions about the world, but self-consciously counterfactual features of models (Ismael 2016, 29; Mäki 2020, 219-220). But the bigger problem is that ‘leaving out’ facts is typically indistinguishable from ‘falsifying’ them (Portides 2021, 5874-5). Does modelling agents as ‘egoistic’, for example, ‘leave out’ non-egoistic motives or make ‘false’ motivational assumptions? To leave any factor out of a model is to treat that factor as if it had zero causal or normative weight, which is usually quite precisely an idealization. Abstraction

\textsuperscript{7} ‘Idealization’ can be a misleading term since it need not model ‘better’ situations. As Johnson (2014, 551-2) points out, for example, Foucault’s ‘panopticon’ is an idealization, but a model of society Foucault considers a warning, not a recommendation.
and idealization are more plausibly understood as orthogonal: abstractness denotes a model’s detail, while idealizations are features of models that are avowedly counterfactual and employed to enable exploration of a model’s distinct (causal, constitutive, normative, or other) properties (A. Levy 2021).

The full implications of these points – that no normative theory can idealize completely, and that all normative theories idealize somewhat – have not been adequately appreciated in the debate over idealization and ideal theory. Thus far, the debate has principally revolved around the question of how ‘the ideal state’ of justice (or some other value) is relevant for normative theorizing in the nonideal world. But the above points entail that there is no singular ideal state of most normative values. We can only ask what justice (or whatever) would require under various kinds of conditions, to which we could always add some further idealization or impose some more realistic constraint. One can theorize justice under conditions of full compliance, for example. But that does not produce the ideal of justice, any more than justice theorized under conditions of maximum moral motivation or resource superabundance or perfect information produces the ideal of justice. Indeed, the specific idealization of full compliance can be used in a model to interrogate any set of principles, ‘ideal’ or otherwise. One might identify ‘ultimate’ principles under full compliance, recognize that there will be non-compliance with those principles, and then design remedial principles to handle such non-compliance. But one can then query how those remedial principles will work if fully complied with – and so on. Whether we analyze full (or any other level of) compliance is therefore separate from the question of whether we root our theories in a model of ‘ideal justice.’

---

8 Even the most sophisticated recent analyses typically present idealization as modelling “an ideal society,” “the ideally just society,” “best-case scenario”, “best feasible scenario”, “ideal institutional design” or “fully just society” (Wiens 2023, 786-7; 2024, 10, 23 & 32; Sirsch 2021, 4 & 13; Chahboun 2019, 551; Barrett 2023, 32) as though these were singular states of normative optimization.

9 On the distinction between principes and models, see: Wiens (2024, 25-9).
Throughout the rest of this paper, I therefore stick to the language of ‘idealization’ rather than ‘ideal theory,’ and when I write of ‘defenders’ and ‘critics’ of idealization, the former refers to those who defend high degrees of idealization and the latter to those who seek to keep idealization relatively minimal. This reframing already generates significant impetus for the heuristic understanding of idealization I will defend. But it does not challenge Rawls’ ‘division of labor’ presentation of more and less idealized theory, nor refute his claim that less idealized theory requires principles derived through more idealized theory. It is that claim (rather than the binary ideal/nonideal distinction), which is subject to the deepest disagreement between defenders and critics of idealization.

Any arguments for or against idealization depend on some interpretation, however tacit, of how its inferential logic is meant to work: i.e. how the depiction of certain idealized circumstances might help us reach certain normative conclusions. Such questions of inferential logic are common across human inquiry, but more familiar in the empirical sciences. We understand ‘controlled comparison’ as important in making causal inferences, for example, because a claim that one variable has causal power over another implies that variation in the first variable will be associated with variation in the second when other factors remain constant.

Political theorists’ understanding of idealization’s inferential logic has, however, typically been left implicit. Its defenders often rest satisfied with clarifying idealization’s purpose, but without explicating how it inferentially serves that purpose. Rawls, for example, offers little explanation of his assertion that: “The reason for beginning with ideal theory is that it provides, I believe, the only basis for the systematic grasp of these more pressing [nonideal] problems” (Rawls 1971, 9). The same can be said of many critics. Sen’s (2006, 221-2) influential claim
that an ideal conception of justice is not necessary to assess more or less just outcomes, any more than identification of the world’s highest point is necessary to determine whether one mountain is taller than another, largely question-begs the inferential logic of idealization. As Schmidtz (2011, 774) observes: “Sen scarcely gestures at an argument here, or even at a clear thesis” (see also: Barrett 2023, 35-7). We must therefore probe deeper to pinpoint the inferential logic of idealization. I argue that two distinct understandings of that logic can be identified.

2. Telic Idealization

The first way of understanding idealization presents idealized scenarios as a necessary way of identifying the ultimate target, end-state, or benchmark for action which further normative principles logically presuppose. Nonideal or less idealized theory is then essentially downstream normative strategy: clarifying what specific actions or institutions we should employ in our actual circumstances to meet the normative standards identified in more idealized theory.  

This telic image of idealization, as I call it, is the dominant portrayal amongst almost all of idealization’s critics, and at least a good proportion of its defenders. Amongst defenders, Simmons (2010, 12) writes: “Where ideal theory dictates the objective, nonideal theory dictates the route to that objective.” Robeyns (2008, 345 & 352) contends that ideal theory provides the “ultimate goal” and “endpoint of our journey” in “the principles of justice in a fully just society.” Approvingly characterizing Rawls’ approach, Sirsch (2021, 2) writes: “in order to

\[\text{There are different versions of telic idealization. Wiens (2015) distinguishes “target” from “benchmark” views, and there is also a difference between those who see idealized scenarios as themselves something we should realise (Robeyns 2008; Uberti 2014), and those who see idealized scenarios as merely revealing our ultimate normative standards or goals (Estlund 2020). My thanks to an anonymous reviewer for highlighting this. For my purposes, what matters is that these stances share the telic inferential logic in which normative judgments presuppose ultimate normative ‘ideals’}.\]
find out what to do here and now, we should use an ideal society as a blueprint.” Amongst the critics, Hope (2016, 387) rejects views in which criteria of justice are “derived from an ideal state of affairs,” Hall (2016, 76) critiques “unconstrained utopianism” for how it involves theorists “outlining their visions of a morally perfect society and using these visions as preliminary blueprints for political design,” while Wiens’ (2024, 10) critique of normative uses of idealization presents them as deriving normative conclusions from “claims that a particular model of society is ideal.”

Why, these critics argue, would idealized models that assume away so many features of our real normative circumstances offer relevant guidance? Surely, as Sen (2010, 18) writes: “justice cannot be indifferent to the lives that people can actually live”? How could it not be the case, Hall (2016, 77) asks, that “we need to decide which model best represents the kind of practice we are seeking guidance about”?

Some such criticisms miss the mark. Defenders of idealization do not claim that we should be indifferent to people’s possible lives or to accurate models of practices in our normative thinking as a whole. Such issues are obviously relevant, they stress, for normative proposals, as Estlund (2020, 11-12) puts it, concerning what policies or institutions to enact in the real world, and ideal theory isn’t intended to generate such proposals on its own (Watson and Hartley 2018, 32; Chahboun 2019, 559; Sirsch 2021, 5; Erman and Möller 2022, 525 & 535). Instead, what defenders of telic idealization contend is that advocacy of any proposal depends on some kind of underlying principle that must be identified at a more idealized step of analysis. Thus Rawls (1999, 90) claims that: “until the ideal is identified…nonideal theory lacks an objective, an aim, by reference to which its queries can be answered.” Simmons (2010, 34)

---

11 Wiens offers a formal model of how telic idealization might work in: Wiens (2023, 786-7).
expresses the same idea in arguing: “The requirement that nonideal policies be ‘likely to be successful’ requires that we know how to measure success; and that measure makes essential reference to the ultimate target, the ideal of perfect justice.”

The inferential logic in this telic image of idealization is thus presuppositional and asymmetric: nonideal normative judgments logically depend, it is argued, on prior judgments about values under more idealized conditions, while those prior judgments do not have a parallel dependency on judgments in less idealized circumstances (see Wiens 2024, 29). This constitutes what Estlund (2020, 7) terms the priority claim: “sound understanding of what is required in realistic nonideal conditions is severely constrained without a prior sound understanding of the requirements under conditions of full compliance.” Or, as Simmons (2010, 34) argues, the fact that proposals are evaluated according to some underlying principles of how society should ideally be operating, “necessitates that ideal theory have priority over nonideal theory.”

These claims seem, however, to jump a logical step. Real-world normative judgments may depend on some underlying principles. But that doesn’t entail that extensive idealization, up to the point of imagining an ideal of ‘perfect justice,’ is necessary to justify such principles. It’s not clear what the move from the first to the second of these claims is intended to be: in general, defenders of idealization assume rather than demonstrate that one implies the other. For example, while rightly explaining the importance of interrogating our underlying principles, Swift (2008) largely equates such analysis with an ideal-theoretic account of “spotless justice.” Simmons (2010, 34-5) and Sirsch (2021, 12-15) argue that without an ideal theory of justice, we cannot know whether apparent short-term improvements take us further along a path to ideal justice or move us onto a different societal path where ideal justice is no longer feasible.
But these arguments rest on the premise that there is a relevant ‘ideal of justice’ that has some warrant to constitute ‘the target’ for improvements in the nonideal world – they do not substantiate that premise. Robeyns (2008, 345), meanwhile, simply stipulates that “reaching Paradise Island [Robeyns’ metaphor for ideal theory] is our ultimate goal. It gives us the direction in which we should be moving to reach a (minimally) just society.” She offers no further justification of that orientation.

Such arguments do not explain why or how accounts of justice (or any other value) in highly idealized settings are needed to provide a relevant ‘target’ or ‘benchmark’ for principles in less idealized settings (see also: Phillips 1985; Pogge 2008; Jubb 2009; Wiens 2023, 786-7; Barrett 2023). Nor, as Wiens (2024) critiques at length, do they show why any specific idealized scenario is the right one for providing such a target or benchmark – remembering, again, that there is no singular ideal state or setting that naturally presents itself for this purpose. Nor do they establish why appealing to highly idealized scenarios is an epistemically reliable or necessary way to justify our principles (Anderson 2009; Jubb 2009; Hall 2016; Sen 2010; J.T. Levy 2016, 327-32; Appiah 2017, 168; Gaus 2016). Indeed, extensive idealization may seem epistemically unreliable. Given the remoteness of highly idealized settings from our actual world, it is not clear that we can confidently determine what principles or institutions would be best in such settings (Elster 2011).  

Thus, defenders of idealization may value consideration of real practices and contexts in normative theory, and their claim that normative judgments in the real world depend on some underlying principles remains plausible. But this does not deflect the charge of the critics: that

---

12 As Estlund (2020, 7) observes, this query was an important part of Henry Sidgwick’s critique of Herbert Spencer in the late 19th Century. For Estlund’s analysis of this problem see: Estlund (2020, 202-3).
a justificatory appeal to ultimate ideals of normative perfection is unnecessary and unhelpful. Perhaps further defenses of telic idealization can address this charge. But I suggest that a more promising move is to explore an alternative way of understanding idealization, that does not suffer from the problems telic idealization generates.

3. Heuristic Idealization

A heuristic understanding of idealization dispenses with the telic idea that normative principles logically presuppose a prior ‘ultimate ideal.’ Instead, it presents idealized models (of varying types) as instrumental tools for “exploring normative scenarios” (Mäki 2020, 220; see also Favara 2023, 387-90). The idea that counterfactual models might be used in some such heuristic manner is widely recognized (Johnson 2014; Dowding 2022). My key claim here is that it is this heuristic usage, rather than a telic conception of idealization as revealing presupposed ultimate ideals, that provides the best way of understanding justificatory appeals to idealizations and ‘ideal theory’ across political theory.

There are various ways idealizations could be heuristically useful. But a principal one – which most directly substitutes for telic idealization’s logic of inference – is as a means for establishing whether our principles are counterfactually coherent. By this I mean that idealized models are used to investigate if the normative implications of our principles shift across different counterfactual settings appropriately according to relevant differences between those counterfactuals.

The inferential claim of the idealizing theorist, here, is not that the justification of a principle derives from its justification in a highly idealized setting. It is that if a principle were justified in any actual scenario, then it will carry implications for various relevant idealized
counterfactuals that should also be justified. Idealization thus serves, not to generate some ultimate warrant for our principles via an ideal of spotless justice, but to help form a system of principles whose varying implications across different counterfactual contexts are well explicated and understood (see also: Estlund 2020, 9; Erman and Möller 2022, 537).

For example, full compliance in Rawlsian ‘ideal theory,’ where principles of justice are formulated in a setting where all individuals will comply with such principles, has baffled many critics when understood in a telic manner (e.g. Schmidtz 2016, 6-7; Rosenberg 2016, 57-60). The critics’ problem is not simply that real-world noncompliance is inevitable and alters what we should do. Rawlsians plausibly reply that, nevertheless, to identify something as noncompliance, we need to know what principles should be complied with (Simmons 2010, 34). The deeper problem is akin to the normative paradox of cognitive angels discussed in Appiah (2017, 91-97), where a cognitive angel is a being capable of perfect rationality but who, because of this, could never encounter many cognitive dilemmas humans actually face. The cognitive angel therefore cannot provide any guidance for how we should respond to such dilemmas: it simply isn’t meaningful to ask what it would do when faced with them (see also: Schmidtz 2016, 9-10). Similarly, the deepest concern with full compliance (or other highly idealized) models is that many normative problems cannot even be posed within such models in a way that yields relevant guidance (Anderson 2009, 132; Wiens 2024).

But there is a more straightforward argument for utilizing models of full compliance for assessing counterfactual coherence. Again, it is not that we derive the right thing to do from idealized conditions of full compliance. It’s that when we have some claim that action X or

---

13 I’m not assuming that normative principles (even ‘ultimate’ principles) must have universal scope (see Pogge’s 2008, 463-9 critique of Cohen 2008 on this). Most principles in political theory are contextual yet not confined to only one setting – see Lægaard (2019).
principle Y is right, we should check whether we would really approve of a world in which everyone did action X or followed principle Y. If not, something further has to be said. Precisely because a world of full compliance is so unlike our actual world, we should not rush to the conclusion that our initial view in favor of action X or principle Y was wrong. But such an idealization suggests that either we were incorrect that X/Y is right, or X/Y is right due to some identifiable aspect of real-world noncompliance.

Idealization in political theory, understood this way, looks less like a curious idiosyncrasy of the discipline, and more like political theory’s version of a more standard aspect of human inquiry: the use of knowingly idealized models (again, not ‘false assumptions’ about the world) that allow us to reveal, test or employ ideas in ways that stretch beyond simply looking at our singular reality (Ismael 2016; Appiah 2017; Nefdt 2021; List and Valentini 2016, 544-5; Dowding 2022). Certain counterfactual idealizations might still be particularly helpful or offer especially powerful forms of justification – because, for example, they show how conclusions we are confident of in certain settings can support more contentious conclusions in other settings. Unlike in telic idealization, however, there is no reason to think that these especially helpful idealizations must be representations of ‘perfect justice’ or ‘the ideal society.’

Why is this pursuit of counterfactual coherence important? To some, the answer will seem obvious. Endorsing a normative judgment in one setting while rejecting it in another without being able to account for how the two settings are relevantly different suggests incoherent thinking that is prima facie unsound. Ad hoc assemblages of principles, with no clear reasons

---

14 This is notably similar to how Rawls interprets Kant, as discussed in Gaus (2016, 22-3). Rawls’ understanding of reflective equilibrium, his focus on the coherence rather than ‘truth’ of ethical stances (Rawls 1975), and his contention that the original position allows a “conjecture” (rather than logically entailed conclusion) as to what conception of political justice its parties would select (Rawls 1999, 30), all might suggest that Rawls’ use of idealizations is more heuristic than telic. But pursuing this exegetical possibility is beyond the scope of my analysis here.
for employing one principle in situation A and another in situation B, are not a systematic form of normative understanding and lack ‘robustness’ in the face of changing conditions (Kirshner and Spinner-Halev 2023). Robust normative theories would allow us to understand how a coherent body of principles interacts with contexts to produce sound normative judgments (Lægaard 2019, 955). This is consistent with prevailing understandings of reflective equilibrium – “the dominant method in moral and political philosophy” (Knight 2023) – where we seek principles which account for our judgments across different contexts (in combination with those contexts’ properties), reciprocally moving back and forth between modifications of our judgments and principles until we have a coherent system of both (Cath 2016, 214-17; Slavny, Spiekermann, Lawford-Smith, and Axelsen 2021). This method is generally understood as justificatory – contra Wiens’ (2024) view that idealizations cannot serve normative justification. Demonstrating the coherent verdicts of a theory across contexts, including idealized counterfactuals, is a part (though probably not the whole) of the justificatory process used to support normative theories, as with empirical theories (Cath 2016, 215-17; Dowding 2022, 517; Knight 2023).

Some political theorists, however, are less inclined to see systematic coherence as valuable. They might, for example, be “particularist contextualists” who deny the existence of general principles that hold across contexts (Lægaard 2019, 956-7). Or they might have a broader skepticism about the value of abstract systematization in ethical thought (e.g. Williams 1985/2006, 219-24; Bourke and Geuss 2009).

Yet there can be value in interrogating counterfactual coherence via idealized models even for such theorists, for at least three reasons. First, such an approach is simply advisable for addressing uncertainty: we should be prepared for unfamiliar normative scenarios and might
want some long-term normative goals that can be pursued even as social circumstances change (Sirsch 2021, 5 & 12-14). This need not require a universalizable theory of ‘the ideal society.’ But establishing that our principles are somewhat robust across different counterfactuals helps avoid fragile ethical frameworks suitable only in settings very close to current conditions (Kirshner and Spinner-Halev 2023). Second, interrogating the counterfactual coherence of our normative principles can help cultivate good normative judgment. Many sceptics of abstract normative theories stress the importance of judgment: understood as a less formal way of using our principles to analyze concrete cases (Beiner 1983, 2-6; Bourke and Geuss 2009). But without idealized counterfactuals, judgment can only work off real-world cases furnished by history, many of which are not especially relevant to the dilemmas that concern us. Counterfactuals can function as an exercise in “practical imagination” (Bourke and Geuss 2009, 6), to address this problem. Finally, interrogating counterfactual coherence might simply be used more critically: to challenge the supposed cross-contextual suitability of a theory or a putatively general principle (Dowding 2022, 517). These three considerations may encourage more realistic idealizations than the pursuit of systematic reflective equilibrium. But they still suggest a need to interrogate counterfactual coherence via heuristic idealizations.

Analyzing counterfactual coherence is not the only heuristic use of idealizations in political theory. Idealizations can also render complex normative scenarios more cognitively manageable (Appiah 2017, 129-35; A. Levy 2021). They might encourage awareness of the possible contingency of existing social arrangements (Appiah 2017, 159-60; Bertea 2023; see also: Stuart 2022). They might help us clarify the conceptual content of normative principles (Wiens 2024). They could have rhetorical and affective functions, that strengthen the pull of certain principles by better engaging our deepest ethical sentiments (Appiah 2017, 152-3). There may be other uses, and idealizations will typically perform multiple functions
simultaneously. I have focused on counterfactual coherence only because it constitutes the clearest substitute for the telic understanding of idealization: in accounting for justificatory appeals to idealizations, but in a way that does not depend on questionable claims about presupposed ultimate ideals.

4. Advantages of Heuristic Idealization

The heuristic use of idealization for investigating counterfactual coherence shares some features with telic idealization. It still stresses the importance of underlying principles that inform judgments in individual contexts, and it continues to rely on the explication of logical connections between judgments in different contexts as a strategy of justification or critique. Yet the differences between telic and heuristic idealization are significant and carry several important implications – generally to the advantage of heuristic idealization – for how we conduct political theory.

First, heuristic idealization escapes a fundamental problem with telic idealization that no one seems to have posed a successful solution to: namely, how to identify the proper idealization threshold (Appiah 2017, 166; Wiens 2024; Hall 2016, 83-87; Rosenberg 2016, 55-6). As noted, rather than a binary ideal/non-ideal theory distinction, idealization comes in degrees. Efforts to address a normative question must assume some fixed background of facts that cannot be changed and some mutable foreground of facts (usually associated with human choices) that can. Rawls, for example, does not idealize everything: he takes things like “the basis of social organization and the laws of human psychology” as facts which institution-designers in the original position treat as given (Rawls 1971, 137). Cohen (2008) critiques Rawls for not

15 It is not committed to the view that such principles are fact-insensitive – see: Pogge (2008); Jubb (2009); Hall (2013); Sangiovanni (2016).
idealizing enough, seeing more human sentiments as mutable and thus potentially unjust. Others think Rawls idealizes too much: treating (almost) inevitable facts about real-world politics as irrelevant to our fundamental principles of justice. How do we determine who is right here? In identifying our ideal goal, end-state or target, what determines the proper degree of idealization? It’s not clear that there is any systematic way of answering this question – different positions, as Rosenberg (2016, 57-60), Hall (2016, 82-91) and Wiens (2024, 16-17) argue, seem arbitrary.

Heuristic idealization evades this problem. There is not some specific level of idealization which is assumed to be the one that defines ‘ideal,’ ‘spotless,’ or ‘perfect’ justice. Instead, normative theories consist of different principles, and those principles will have different implications in different (more or less idealized) scenarios. What matters is not to find the idealization to generate our ‘ideals,’ but to utilize various appropriate counterfactual idealizations to build and interrogate a theory of our principles that can explicate and vindicate the prescriptions they yield in different settings (see also: Favara 2023, 389-91).

Take, for example, Rawls’ (1999, 23) claim that the basic social unit for analyzing ideal justice in international politics is “peoples acting through their governments.” In telic idealization, this claim is inferentially unstable. Why not, alternatively, see some kind of world government as ideal? Because, says Rawls (1999, 36), a world government would likely constitute a global despotism or a fragile empire. But, while that is an eminently sensible concern in our actual world, we can theorize an ideal world government which would not suffer from such problems. Rawls is building factual assumptions about real-world politics into his ideal theory here. Conversely, why not start with an ideal of a world where governments have no privileged status, and all individuals and groups can seamlessly form their own laws and punish
wrongdoers – perhaps even waging private wars, as argued in Fabre (2008)? If our normative judgments are taken to rest on some telic ideal, it is hard to see how we adjudicate between these various idealized possibilities.

With heuristic idealization the inferential logic proceeds differently. We can start with any putative principle (though it may make sense to start with one which seems potentially applicable to the actual dilemmas we face). For example, we might hold the principle ‘only peoples acting through their governments may wage war,’\textsuperscript{16} on the grounds that governments exist to protect and promote the rights and interests of their peoples. This appeals to a somewhat idealized model of government – not all governments are like this – yet one rooted in our real world – this is how we perceive most governments. It can be challenged, from both other idealizations and the real world. What about the alternative claim that only a world government should wage war, justified through an idealized model of world government as an effective, authoritative, and impartial representative of all humanity? If such a world government existed, we might concede, perhaps only it should employ war. What about national liberation movements fighting against repressive colonial governments? Where such movements effectively promote the rights and interests of their peoples, we might agree, they might justifiably fight wars. Our original principle seems counterfactually incoherent. We might therefore revise it: ‘wars may only be fought by those available political institutions which best protect and promote the rights and interests of their peoples.’

There would obviously be much more to say. Such a principle may look justified when fully complied with yet prove destabilizing in practice: encouraging lots of non-governmental actors

\textsuperscript{16}Obviously, this principle would be only part of a set of broader requirements, such as those of Just War Theory, collectively specifying when war is justified.
to wrongly claim authority to wage war. This might suggest further modifications of the principle or the need to balance it against other principles. The method here is to engage in a comparative analysis across a range of different counterfactual (and real-world) cases to gradually refine our overarching normative theory. We thereby avoid the twin dangers of our principles either being too restricted by contingent features of the immediate real-world scenarios we face or being valid only in idealized conditions fundamentally unlike the real world (Valentini 2012, 659-60).

A second advantage of heuristic idealization is that it renders the question of what makes any possible idealization appropriate more tractable. For a start, this question becomes problem-relative. Rather than identifying the ideal model of society, we are constructing whatever idealizations most effectively assess potential inconsistencies in, or build cross-contextual support for, our normative theories. This depends on the nature of those theories, their subject matter, and their most likely vulnerabilities. It will also depend on a political theorist’s metaethical and methodological orientation: idealized counterfactuals could be pure intuition-pumps for philosophical reflection, cases for the application of axioms, the basis for laboratory experiments to test real-world actors’ attitudes and behavior, settings for formally modelling the implications of certain normative assumptions, or much else besides. I remain neutral, here, between these broader methodological strategies.

But heuristic idealization clarifies why certain employments of idealization are wholly inappropriate in normative argument. I build here on Valentini’s (2009, 353) contention that what makes certain idealizations “bad” is that “facts are not ‘innocently’ denied at the level of theory construction, but their denial is also presupposed as a condition for the applicability of the theories themselves.” While Valentini illustrates this idea with examples from the work of
Rawls, Kymlicka and Dworkin, she does not fully explicate what differentiates these two kinds of idealization, i.e., how to tell whether facts are being denied in theory construction or as a condition for theory application. As Erman and Möller (2022, 538) observe, such assessments are frequently “not as clear-cut as it seems” – the idealization remains justificatorily in play at both levels. My suggestion is that the “innocent” denial of facts in theory construction that Valentini describes is not really denial at all, because there’s no assumption of a falsehood in the normative theory, merely the modelling of a counterfactual for heuristic purposes. By contrast, when denial is “presupposed as a condition for the applicability of the theories themselves,” action-guiding principles are defensible only if we actually make false assumptions about the context in which those principles are applied.\footnote{This reconstruction seems consistent with Valentini’s later comments in List and Valentini (2016, 545).} That is a good reason to think that such principles are false, irrelevant, or need to be reformulated for the real context we are in.

Heuristic idealization therefore cautions against both the categorical dismissal of ‘distant’ idealized models (i.e. ones very unlike our actual world) and the presumptive privileging of such models. Instead, there is typically a trade-off, analogous to the distinction in the natural and social sciences between an experiment’s “internal” and “external” validity (Dowding 2022). When effectively constructed, more distant idealizations may gain some internal validity. By being more stylized and specified, they enable more controlled normative comparisons of different counterfactuals, removing contingent features of the real world that might confound our judgments so that we may probe the implications of particular normative features more clearly (see Brownlee and Stemplowska 2017). They might thereby also help us identify ambitious or long-term targets for political action. This is not because they reveal ‘the
ideal society’ or ‘peak of justice’, but simply because they remove confounding features of our actual world that we have reason to think are contingent, so that we can analyze how our principles would play out in worlds representing significant apparent improvements on our own.\textsuperscript{18}

At the same time, however, distant idealized models may lose some ‘external validity.’ Precisely because of their stylization and distance from more realistic cases, it may be harder to extrapolate from highly idealized models to arrive at general conclusions for other (including real-world) settings. Indeed, it is crucial to appreciate that overly-fanciful idealizations can also lose internal validity: by being so unlike our real world that we lack enough information to properly specify or interpret their own features and the normative judgements they would imply (Dowding 2022, 519; Elster 2011).

For these reasons, a heuristic understanding of idealization will tend to encourage less distant idealizations by comparison with a telic understanding. It also does not privilege idealized models over consideration of real-world cases. As noted, idealized models provide an important capacity to probe beyond the limits of historical and present reality. But examining actual political scenarios can also interrogate counterfactual coherence. For this reason, heuristic idealization is more compatible with approaches like political realism (Sleat 2018), grounded normative theory (Ackerly et al. 2024), and ethnographic political theory (Longo and Zacka 2019), which stress the need to situate theorizing in reflection on real political settings. Nevertheless, a heuristic understanding can also handle the use of distant idealizations to think

\textsuperscript{18} As observed in footnote 7, however, idealizations do not necessarily model counterfactuals ‘better’ than present conditions. A principal interest in normative progress may warrant greater focus on such positive heuristic idealizations, but political theorists (especially those more pessimistically inclined) may also be interested in assessing principles in counterfactuals worse than present conditions (Barrett 2023, 50). My thanks to an anonymous reviewer for pressing me on this point.
about ambitious normative goals in the way many defenders of idealization desire (Bertea 2023).

Third, as this suggests, heuristic idealization is more methodologically ecumenical: offering a stronger case for diverse approaches in political theory to value idealization, including those of most critics of telic idealization. Like Mills (2005), for example, one may be centrally interested in justice claims arising from the real-world history of racial oppression and exploitation. One may further share the assumption of those like Mills, Haslanger (2012, 24-30), Ackerly et al. (2024) and others, drawing on standpoint epistemology, that to even meaningfully comprehend racial oppression and exploitation one must engage deeply with the lived experiences of those subject to it rather than concentrating on abstract ideals. Like political realists and practice-dependent theorists, such as Hall (2013), Jubb (2009) and Sangiovanni (2016), one might think that our principles surrounding a phenomenon like exploitation depend on factual features of the practices they are used to evaluate. One may share Sen’s (2010) view that our normative principles surrounding racism and exploitation are primarily ‘comparative’, and manifest in examination of different real-world cases. Even accepting all this, one must still recognize that identifying any actual instance of ‘exploitation’ depends on some normative standards, which will carry counterfactual implications about how our relations would be less exploitative if we treated each other differently (Estlund 2020, 17-18). An idealized model of the world in which we engaged in less such apparent exploitation thus serves as a tool for revealing and interrogating such putative normative standards and their implications.

Again, the claim here is not that if the behavior we see as exploitative in the real world doesn’t seem exploitative in idealized conditions then we were wrong to think it exploitative in the real
world. What such a test would imply is that either we are wrong to see such behavior as exploitative or that the behavior is exploitative in virtue of some distinctive normative context of the actual world or that some alternative principle(s) or application of our principle(s) is necessary to render our judgements coherent across the two scenarios. These can be important conclusions. Thinking, for example, that selling one’s body for sex is intrinsically exploitative (and should not occur even in highly idealized circumstances) and thinking that it is exploitative and unjustifiable only in virtue of certain structures of patriarchy and unequal power in actual world settings, are distinct normative positions, with distinct real-world implications.

Given these advantages, my contention is that most political theorists, across different methodological orientations, can accept a heuristic understanding of idealization and its importance. There will be exceptions. Some scholars inclined to telic idealization could object that, while heuristic idealization may appear less contentious, it involves a considerable contraction in ambition, since it cannot generate the ultimate warrant for our principles that telic idealization offers. Put another way, one might think that heuristic idealization simply substitutes an ambitious foundationalism about normative knowledge with a more modest coherentism.

This isn’t quite right. Heuristic idealization is neutral on the question of ultimate normative foundations: maybe there are none, or maybe they lie in bedrock intuitions, immanent consequences of practices, local intersubjective norms, transcendental truths, or whatever. Whether our principles have such foundations or not, we want to understand the principles coherently – and interrogating how we apply and evaluate them across counterfactuals is one crucial way to do that (see also: Gaus 2016, 26-9; Cath 2016, 218-20; Knight 2023). Nor does
telic idealization have to be foundationalist – even Cohen (2003, 222-4), one of the strongest defenders of highly idealized principles, contended that his approach could be compatible with coherentism.

Still, heuristic idealization may still seem too indeterminate for some scholars (parallel concerns exist with reflective equilibrium, see: Cath 2016, 220-8). Assessing counterfactual coherence cannot itself provide the kinds of ultimate foundations for normative principles that telic idealization may aspire to, nor adjudicate between multiple equally counterfactually coherent theories (although such deficiencies might be compensated by also employing other methodological strategies – heuristic idealization is only one tool among others). Conversely, some political theorists may maintain that idealized models are going to be so unreliable and dependent on dubious prior intuitions that we should avoid them as far as possible in favor of a focus on historical knowledge, concrete practices, and real-world dilemmas.

There will still be divergences, moreover, even among those who do accept heuristic idealization as an important part of political theory’s methodological toolkit. Theorists will likely differ, for example, over quite how methodologically central heuristic idealizations should be. But rather than reflecting deep disputes over how to build normative arguments, these differences are likely to reflect contrasting but compatible vocational priorities. For scholars like Sen and Mills, the truly urgent problems involve understanding and tackling manifest injustice in the world around us – injustice so manifest that we do not need to spend much time testing our underlying standards involved in deeming it unjust (see also: Mills 2015, xi-xv). For scholars like Ismael (2016, 30), on the other hand, “our pre-theoretic conceptions of injustice are too thin a reed” to have strong faith in, such that clarification of our principles is a crucial task (as Swift 2008 also argues). Nothing seems wrong with political theorists
dividing their attention this way. What becomes crucial is that they are clear about their goals and *exactly how* particular heuristic idealizations serve those goals. Similarly, the trade-off in the validity of distant counterfactuals is likely to be one that political theorists balance differently: with some continuing to value distant counterfactuals as a way to construct very focused tests of our judgments, while others think them too distant to produce reliable insights. There is further debate to be had here. But these differences can constitute complementary intellectual projects built on a shared understanding of idealization, rather than the seemingly deadlocked present antagonism between defenders and critics of ‘ideal theory.’

**Conclusion**

Many practitioners and most critics of idealization have depicted it as a way of revealing the ‘ideal’ form of justice (or some other value), from which other normative judgments are purportedly derived. I have argued that this is neither a necessary nor compelling way of understanding idealization. Instead, idealization is better understood as a heuristic device that political theorists employ for analyzing principles, values and/or standards, wherever they derive from. It is, amongst other things, crucial for assessing and enhancing the counterfactual coherence of our normative theories and the resulting robustness of our principles across different contexts. This approach involves no contentious claims about the priority or logical presupposition of ultimate ideals. It aims at a more uncontentiously valuable kind of theory: one that provides the clearest possible account of our principles and how they apply and yield different plausible implications across a range of possible circumstances.

There is a more general implication to this reframing of idealization: namely, that the way we talk about ‘ideals’ in politics (both in scholarship and ‘real politics’) is often misleading. The kind of talk I am referring to centers on a common way of construing a distinction between
‘idealists’ and ‘realists’ (or sometimes ‘pragmatists’), as though this is about how much the two camps rely on ideals. It is a construal that has appeal for both camps: idealists condemn realists for their conservatism and lack of principles, realists condemn idealists for their ineffectual utopianism. Yet once we understand idealization as inescapable and heuristic, this construal is exposed as a conceit. All normative thinking revolves around principles which carry implications under both the conditions we face right now and counterfactual conditions unlike our immediate context. That is true for both idealists and realists. What is really going on in (objectionably) strong forms of idealism is not the maintenance of purer normative principles called ‘ideals,’ but a refusal to consider variation in the plausible implications of those principles across different practical contexts. What is really going on in (objectionably) strong forms of realism is not a hard-headed prioritization of ‘what works,’ but a refusal to consider whether one’s judgments as to what counts as something working are really ones which can be coherently justified across different cases.

In this sense, while heuristic idealization does not directly resolve the crucial debate of how normatively ambitious – how ‘idealistic’ – we ought to be in real-world politics, it rightly frames this debate as a substantive ethical and practical disagreement, not a matter of methodology in which one side can just be dismissed as committing some kind of technical mistake. Both radical revolutionaries and cautious reformers make idealizations: what divides them are different underlying principles, different judgments about justice and injustice in the status quo, and different estimates of what is feasible.

Such disagreements are irrevocable parts of the substance of politics. Scholars, citizens and politicians will always dispute how far we should look ahead, how achievable specific policies or goals are, how many risks we should run to pursue them, and to what extent we should
accommodate, coerce or seek to persuade those who stand in their way. Political theory needs to highlight, clarify, and address such substantive disagreements, not bury them under methodology. By contrast, I have sought to demonstrate that most methodological disputes over ideals and idealizations that have dominated political theory since Rawls may be essentially dissoluble. We can get past them: and recognize idealizations as just one of the many essential tools that political theorists – like other researchers – use to analyze our actual and possible worlds.
References


