**Conceptual Ethics and The Categories
of "Ideal Theory" and "Non-Ideal Theory" in Political Philosophy:
A Proposal for Abandonment**

Tristram McPherson (Ohio State) mcpherson.164@osu.edu
and
David Plunkett (Dartmouth) david.plunkett@dartmouth.edu

Introduction

In contemporary political philosophy, it is common to distinguish "ideal" from "non-ideal" theory, and to explain projects or arguments in terms of this distinction. At the same time, there is a burgeoning literature in which philosophers argue about the relative merits of engaging in "ideal theory" vs. "non-ideal theory". Where one stands with respect to this debate is often treated as a central methodological faultline in political philosophy and other parts of social and political inquiry. Further, the language of "ideal" and "non-ideal" is infectious, with philosophers now discussing "non-ideal" theories in ethics, epistemology, and the philosophy of language.[1]

      This paper makes a prima facie case for *abandoning* use of the terms 'ideal theory' and 'non-ideal theory' in social and political inquiry (across a central range of contexts).[2] Our argument begins by observing two sorts of striking variation. The first is variation in how inquirers characterize "ideal" and "non-ideal" theory. The second is variation in the theoretical significance of

---

[1] See, for example, the papers collected in (Tessman, 2009). See (Cappelen & Dever, 2021) for critical discussion of recent work on "ideal" vs. "non-ideal" theory beyond political philosophy, with a focus on the philosophy of language.

[2] In this chapter, we use italics for rhetorical stress, single quotes to mention linguistic expressions, and double quotes for quoting other authors, "scare quoting", and other informal uses (prominently, simultaneous use and mention).

idealization. Here, there is striking variation across theoretical contexts in the targets that can be usefully idealized, the degree of idealization that can be applied to those targets, and the ways that idealizations can be warranted or worrisome. Against the background of this variety, we argue both that central uses of these terms have the potential to obstruct inquiry, and that there is no theoretically appealing way to reform the existing usage.

It is worth emphasizing that in advocating for abandoning "ideal theory" *talk*, as we do here, we are *not* arguing for a "non-ideal" approach to political inquiry. Rather, as we emphasize in the conclusions, we are arguing that inquirers on all sides of contemporary debates over "ideal theory" would do well to reframe their debates in ways that do not treat "ideal theory" as a central locus of discussion.

We proceed as follows. After more carefully framing our discussion (§1), we present our positive argument, in four stages. First, we introduce two sorts of data: about how the term 'ideal theory' and its cognates are used in contemporary political inquiry (§2) and about the heterogeneous significance of idealizations across theoretical contexts (§3). We then employ this data as the basis for arguing against continuing to use the relevant terms (§4), and against seeking to reform those uses to ameliorate the vices of the existing patterns of use (§5). We conclude by explaining what we take to be the relevant virtues of political inquiry that has abandoned the use of these terms.

1. Preliminaries: conceptual ethics, abandonment, and political inquiry

This section more carefully introduces our topic, and several crucial assumptions that will guide our discussion. In short, our aim is to evaluate the use of the term 'ideal theory' and its cognates in the context of political inquiry. We begin by briefly introducing how we understand political inquiry. We then introduce the broad sort of project we are engaged in, relative to political inquiry: conceptual ethics. Finally, we introduce a crucial assumption about what sorts of aims we will take to be relevant to evaluating conceptual ethics proposals for terminology within normative political inquiry.

As we understand it, political inquiry is a kind of *normative inquiry*. (Here and elsewhere in this paper, we use 'normative' broadly, to encompass the evaluative, the deontic, and the aretaic, as well as the normative more narrowly construed). On the view that we favor, normative inquiry in general aims to answer extensional and explanatory questions about the instantiation of different normative conditions.[3] When normative inquiry concerning politics is conducted at a high level of generality we are engaged in *systematic* normative political inquiry. For example, when we ask what distributive justice consists in, or what it takes for a state to be legitimate, we are engaged in systematic normative political inquiry. By contrast, when we ask normative political questions tied to certain salient clusters of contexts, we are engaged in *applied* normative political inquiry. For example, if we ask about what rights workers should have in the context of a capitalist society in the 21st century, or whether reparations are due to descendants of enslaved people in America given the specific facts of American slavery, we are engaged in applied normative political inquiry. In this paper, we want to evaluate the use of the term 'ideal theory' (and its cognates) in the context of systematic and applied normative political inquiry. To be clear, we take 'non-ideal theory' to be a cognate of 'ideal theory'. So, in making this argument we are as much targeting the use of 'non-ideal theory' as we are targeting the use of 'ideal theory'.

Our argument about 'ideal theory' and its cognates is an instance of *conceptual ethics*. In general, as we understand it, conceptual ethics concerns certain kinds of normative and evaluative questions about our thought and talk, such as questions about which concepts we should use (and why), or which concepts are better or worse (and why). As we understand it, conceptual ethics addresses such questions about concepts, words, or other "representational" or "inferential" devices. Furthermore, it is not limited to the moral or political evaluation of concepts or words, but rather can involve any number of

---

[3] See (McPherson & Plunkett, 2017) and (McPherson & Plunkett, 2021a).

standards for assessing concepts and our use of them (e.g., whether the concepts carve at the joints of reality or not).[4]

Our core question in this paper concerns the evaluation of *language*, not thought. We evaluate the use of the *terms* 'ideal theory' and 'non-ideal theory', *not* the use of the concepts that these terms may be used to communicate. Part of the reason for this focus is that, as we will argue, it is not clear what concepts (if any) these terms pick out.

At a broad level, we can distinguish three competing conclusions that we might draw when engaged in conceptual ethics work on a given term. First, we could conclude that we ought to *retain* the term, with its current meaning and associated patterns of use. Second, we could conclude that we ought to keep the term, but *reform* its meaning and/or use. Or third, we could conclude that we ought to *abandon* use of the relevant term.[5] Our argument in this paper will proceed comparatively, by arguing that abandonment is preferable to retention or reform, in the case of 'ideal theory' and its cognates.

Proposals to abandon a piece of terminology are often coupled with proposals to replace that terminology with superior terminology. For example, Kevin Scharp's work on 'true' and its cognates advocates for abandoning use of this term in certain theoretical contexts, and instead using the novel technical terms 'ascending truth' and 'descending truth', as appropriate, in these

---

[4] Our gloss of conceptual ethics draws from (Burgess & Plunkett, 2013a), (Burgess & Plunkett, 2013b), and (Cappelen & Plunkett, 2020). On our way of thinking about things, conceptual ethics is closely connected to "conceptual engineering". Put roughly, as we understand it, paradigm instances of work in conceptual engineering draw on or involve work in conceptual ethics, in combination with "conceptual innovation", which involves reforming existing concepts (or words etc.), or creating new ones, as well as work on "conceptual implementation", which involves attempts to get relevant agents to actually take up the proposed conceptual/linguistic changes. This basic take on what conceptual engineering draws from (Cappelen & Plunkett, 2020) and (Burgess & Plunkett, 2020). The accounts offered in those two papers differ from each other in some of the details. But those details don't matter here. For some other recent takes on what conceptual engineering involves (and associated discussion of conceptual ethics) see the papers collected in (Burgess et al., 2020), as well as (Isaac, 2020), (Pinder, 2021), and (Chalmers, 2020).
[5] Our use of "abandonment" here draws from (Cappelen, 2023). Note that on some uses of 'eliminativism', to be an "eliminativist about x" is essentially the same as what we are calling "abandoning" either the term 'x' or a concept tied to our use of 'x'. However, in other contexts, 'eliminativism about X' denotes a nihilist thesis in metaphysics: roughly, the denial that Xs exist. It is important to distinguish the nihilist metaphysical thesis from the conceptual ethics abandonment thesis, as emphasized in (Burgess & Plunkett, 2013b). For ease of exposition in this paper, we stick with the terminology of "abandonment" rather than "eliminativism".

contexts.[6] In this paper, we will not advocate for introducing novel "replacement" vocabulary. Instead, in the conclusion we will suggest that we can use other existing vocabulary to clearly focus attention on the virtues and vices of particular candidate idealizations in particular theoretical contexts.

A central question for any conceptual ethics project is: what normative or evaluative standards should we use in assessing the use of the target words or concepts? This is, in general, a deep and difficult question.[7] Here we tentatively propose an approach to answering it, in our context. This approach begins by assuming that one important way to evaluate certain words and concepts, as they are used within a particular branch of inquiry, is by reference to the success conditions (or "aims") of that inquiry.

For simplicity's sake, we will make the further assumption that the success conditions of normative and applied political inquiry are *alethic* or *epistemic*. That is, a normative political inquiry is successful when, and to the extent that, it results in things like true belief, knowledge, or understanding concerning its focal topic(s). Thus, when evaluating our target terms, we will be interested in how well their use serves the project of coming to have true beliefs, knowledge, or understanding of important normative or applied political topics.

This approach is admittedly controversial: some political inquirers appear to think that at least some political inquiry has directly *practical* aims rather than alethic or epistemic ones.[8] For example, perhaps one might hope that political inquiry could help to guide our political lives together, or to promote justice – and that one of these goals might itself set the success-conditions for the inquiry as such.[9] We have focused on broadly epistemic aims for three reasons. First, we think these aims are plausibly important to a wide

---

[6] See (Scharp, 2013).

[7] For discussion, see (Burgess & Plunkett, 2013b) and (McPherson & Plunkett, 2022).

[8] Consider another non-epistemic role for usage of this terminology: saying "I am a [non-] ideal theorist" might function to signal one's allegiance to a broad camp within the academic community, in the same way that saying "I am an *analytic* philosopher" might. We think this sort of signaling usage is sociologically important. However, it is not our focus in this paper, and we thus set it aside in what follows.

[9] For some versions of this idea, see (Haslanger, 2000), (Geuss, 1981), and (Jaggar & Tobin, 2013). For discussion, see (Plunkett, 2016).

range of political inquirers. To underscore this point, one might well think that engaging in political inquiry (where it is understood to have purely epistemic success conditions) could be instrumentally helpful for achieving other, practical goals (such as the promotion of a more just society), even if those practical goals aren't part of the aims of inquiry as such. Second, it's not feasible, given the length of this paper, to develop our argument relative to the full range of possible aims for political inquiry. And, finally, we are optimistic that the conclusions that we draw could be generalized to apply to other conceptions of political inquiry, on which it has non-epistemic aims.

We take it that epistemic or alethic aims can provide a plausible and entirely familiar rationale for the abandonment of words or concepts, especially for theoretical words or concepts. Consider, for example, terms like 'phlogiston' (in chemistry) or 'hysteria' (in clinical psychology). In both cases, abandonment seems decisively warranted, due to the fact that use of these terms has turned out to be counterproductive for investigating the relevant parts of reality.[10]

With this take on the aims of inquiry in hand, we can sharpen our central conceptual ethics question about the term 'ideal theory' and its cognates, in contemporary political inquiry. Given our epistemic aims, is it best to retain these terms with their current meanings, to reform the use and meanings currently associated with these terms, or to simply abandon their use? This paper argues that contemporary political inquirers would (in general) be better off relative to epistemic aims if they abandon use of the term 'ideal theory' and its cognates. Note that this thesis is a generic claim, not a universal claim. Like most plausible abandonment theses, it is compatible with the possibility that there are some (unusual) contexts where it would be epistemically useful for those engaged in political inquiry to continue to use this terminology. Furthermore, it is worth underscoring that since our focus here is on abandoning the use the terminology of 'ideal theory' and its cognates in the context of engaging in political inquiry, our main argument is compatible

---

[10] Note that the case of 'hysteria' helps to illustrate that this sort of epistemic standard is not the only standard relevant for conceptual ethics. Another excellent reason to abandon 'hysteria' was that use of this term functioned to reinforce sexist norms and practices.

with not abandoning that terminology in other contexts (e.g., perhaps in certain contexts when engaged in intellectual history).[11]

With these clarifications in hand, the next two sections prepare our case for abandonment by providing two sorts of relevant data: data about the heterogeneity of the contemporary *use* of 'ideal theory' and its cognates, and data about the heterogeneous *functions* that idealizations can play in political inquiry.

2. The heterogeneous use data

'Ideal theory' and its cognates were arguably introduced as technical terms in political inquiry by John Rawls's *A Theory of Justice*. For Rawls, engaging in "ideal theory" about justice involves asking what just social/political institutions would be like given (among other things) certain *idealizing assumptions*.[12] Rawls appears to make two of these assumptions definitional of "ideal theory". The first is the assumption of full compliance with just institutions by the agents who are regulated by those institutions.[13] The second is that we are to consider a society in circumstances favorable to the establishment of justice.[14] Rawls also makes other idealizing assumptions in *A Theory of Justice*, although he does not appear to make them definitional of "ideal theory". For example, he assumes that agents in the society he describes are not motivated by envy in the way actual humans often are.[15]

Given that 'ideal theory' and its cognates were introduced as technical terms, it is a striking fact that contemporary political inquiry has not tended to

---

[11] We suspect that, if our main argument in this paper is right, then, for many contexts in discussing intellectual history, it will be better to mention (rather than use) the terminology of 'ideal theory' and its cognates. But we don't take a stand on that in this paper.
[12] We do not also offer a gloss on Rawls' characterization of 'non-ideal' theory, because we take the text here to be less clear. (Simmons, 2010) reads Rawls as offering a comparatively narrow account of "non-ideal theory", where, roughly, it concerns the steps we can take on the path towards ideal justice. Some passages in Rawls, however, suggest a broader characterization of "non-ideal theory", including Rawls' claim that such theory concerns "which principles to adopt under less happy circumstances" (Rawls, 1971/1999, 216).
[13] (Rawls, 1971/1999, e.g. at 8; 216.).
[14] (Rawls, 1971/1999, 216).
[15] (Rawls, 1971/1999, §25).

simply defer to Rawls' characterizations. Instead, the contemporary landscape is rife with seemingly cross-cutting characterizations. Consider representative framings from three recent surveys of this literature:

- "Recent debates have seen the term used in different ways and its critics attacking a variety of different targets."[16]
- "A quick glance at what falls under the heading 'ideal/non-ideal theory', however, reveals the heterogeneity of this debate."[17]
- "The notion of an ideal theory can have at least two different meanings.... First, a political theory can be ideal in that it provides an ideal, or goal that we should aim for.... Second, a political theory can be ideal in that it involves models intended to represent certain phenomena, with the associated false assumptions (either abstractions or idealizations)."[18]

To get a feel for the heterogeneity of the use of these terms, consider Laura Valentini's influential attempt at regimentation. Valentini suggests that there are "at least three distinct meanings given to the adjectives 'ideal' and 'non-ideal'" in this debate.[19] The first is the Rawls-inspired idea that ideal theory involves an assumption of *full compliance*. The second is the idea that an "ideal theory" is "utopian" rather than "realistic". This means, roughly, that it is insensitive to hypotheses about feasibility. Finally, there is the idea that an "ideal theory" gives us an "end state" or goal to work towards, whereas "non-ideal theory" is a theory of how to transition towards that goal.

Even Valentini's expansive regimentation, however, fails to capture the full range of prominent characterizations in the literature. Consider some examples.

First, Zofia Stemplowska offers a characterization of the distinction between "ideal" and "non-ideal" theory as concerning whether a theory offers *real-world guidance*. She argues that "one helpful way of understanding the cut between ideal and nonideal theory is roughly to classify as nonideal any theory

---

[16] (Stemplowska & Swift, 2012, 373).
[17] (Valentini, 2012, 654).
[18] (Thomson, 2020, 2).
[19] (Valentini, 2012, 654).

that issues recommendations that are desirable and achievable for us given where we find ourselves and, as ideal theory, any other type of normative theory."[20] At first blush, Stemplowska's distinction does not fall neatly into any of Valentini's broad categories. For example, a "realistic" theory could refrain from offering guidance, while one could offer guidance based on a "utopian" theory. And a familiar thought in the contemporary philosophical literature is that one could offer practical guidance that is not directed towards getting us closer to some "ideal" end state.

Second, consider Elizabeth Anderson's characterization of the distinction. Anderson writes: "In ideal theory, ideals…. are not subject to testing in practice because they set standards, outside of practice, for the success of practice."[21] Anderson claims that nonideal theory contrasts with this as follows: "In nonideal theory, ideals embody imagined solutions to identified problems in a society. They function as hypotheses, to be tested in experience."[22] Anderson's distinction concerns the *epistemological status* of ideals relative to practice and experience. It thus seems to crosscut both Valentini's broad categories and Stemplowska's distinction.

Finally, prominent discussions disagree about whether it is appropriate to classify certain debates as "ideal theory" debates. Consider two significant examples. First, some classify G. A. Cohen's insistence that political inquiry should bottom out in "fact-insensitive" normative principles as a kind of "ideal theory".[23] Others, however, take these claims from Cohen to be orthogonal to the "ideal/non-ideal" cut.[24] Second, while (as we have seen) Valentini treats sensitivity to feasibility as one of her three central characterizations of "the" "ideal/non-ideal" cut, Christopher Thompson suggests that this is a related but distinct issue.[25]

---

[20] (Stemplowska, 2008, 339).
[21] (Anderson, 2010, 6).
[22] (Anderson, 2010, 6).
[23] For example, see (Valentini, 2012, §2.1) and (Thomson, 2020, 2), discussing (Cohen, 2003) and (Cohen, 2008).
[24] For example, see (Enoch, 2018, 11n.11). It is also natural to read (Stemplowska & Swift, 2012, §4) as endorsing the view that Cohen's fact-insensitivity thesis crosscuts the "ideal/non-ideal" distinction.
[25] For example, see (Thomson, 2020, 12).

A further point is worth emphasizing here: on many (though not all) prominent ways of drawing the distinction between "ideal" and "non-ideal" theory, "non-ideal theory" marks out something more specific than just the negation of "ideal theory". For example, consider Anderson's epistemological way of drawing the distinction. On her view, "non-ideal theory" involves a *specific* kind of epistemological alternative to what she takes to be the one involved in "ideal theory", rather than encompassing the totality of possible alternatives. Or, to take another example, consider Tommie Shelby's gloss of the distinction between "ideal" and "non-ideal" theory. In short, he claims that a) ideal theory aims to provide "a comprehensive account of the principles a society must satisfy to be fully just"[26] and b) non-ideal theory "specifies and justifies the principles that should guide our responses to injustices".[27] The latter isn't just *any* kind of theory that isn't the former. Rather, it's a specific kind of other theory, which, on Shelby's view, is meant to complement the former in an overall theory of justice.[28]

This survey is far from exhaustive.[29] But it allows us to draw a striking conclusion. Careful observers are broadly united in concluding that there are multiple substantially distinct prominent uses of 'ideal theory' and its cognates in recent political inquiry. However, they also disagree about (i) how to regiment that range of uses, and (ii) which sorts of views even fall within the scope of discussions of "ideal vs. non-ideal theories". The diversity is significant enough to these commentators that, as we have seen, two of the three recent surveys of this literature we discuss above (Valentini and Thompson) seemingly suggest that the relevant terms are *polysemous*, meaning that they have multiple circulating meanings. The authors of the third (Stemplowska and Swift) frame their discussion by describing the recent literature as marked by

---

[26] (Shelby, 2016, 11).
[27] (Shelby, 2016, 11).
[28] As he puts it: "Ideal theory and nonideal theory are complementary components of an endeavor to devise a systematic account of social justice." (Shelby, 2016, 11).
[29] In addition to not canvassing every view about how to distinguish "ideal" from "non-ideal" theory, we haven't introduced other important aspects of the debate, such as discussion of the primary *purpose(s)* of each kind of theory. For some recent discussion here, see (Shelby, 2016), (McKean, 2017), and (Stemplowska, 2017).

"confusion", "ambiguities", "talking past", and "cross-cutting complexities".[30] David Enoch sums up the state of play as follows: "The terms "ideal theory" and "non-ideal theory" have unhelpfully come to stand for several different things".[31]

3. The heterogeneous function data

We have just introduced the remarkable heterogeneity of the ways the term 'ideal theory' and its cognates are used. This heterogeneity is compatible with the hypothesis that there is a singular way we *could* use this term, which would be especially theoretically valuable. This section introduces reasons to doubt this hypothesis. We suggest that there is not a single determinate thing that it would be useful for 'ideal theory' to pick out. We make our case by focusing on the central notion of idealization. We sketch four relevant sorts of theoretically important variation with respect to idealization in political inquiry.[32] Such idealizations vary with respect to:

- what is idealized;
- the degree of idealization;
- the value served by idealization; and
- the potential liabilities of idealization

We will consider these dimensions in turn.

First, as we have briefly illustrated in the preceding section, political inquirers can and do idealize in a wide variety of ways. For example, they sometimes assume *full compliance* with political norms, or that political agents are *rational*, or *self-interested*, or *informed about politically relevant facts* or *not prone to envy*, etc. They assume that the polity does not have a politically non-ideal *history* with which it needs to wrestle, and that it does not exist in a world that poses an *existential threat* to polities with certain political structures.[33]

---

[30] (Stemplowska & Swift, 2012, 373-374).
[31] (Enoch, 2018, 11).
[32] (O'Neill, 1996, Ch. 2) helpfully distinguishes *idealization* (which, on her account, involves theorizing using claims that are not true) from *abstraction* (which, on her account, involves ignoring certain true claims). We take much of what we say in this section to be adaptable, mutatis mutandis, to abstraction.
[33] Versions of each of these assumptions are made in (Rawls, 1971/1999) and much of the literature following and engaging with that text.

Further, these *actual* idealizing assumptions are only the tip of an iceberg of *potential* idealizing assumptions that it might be useful to make in political inquiry. Call this point the *variety of dimensions of possible idealization*.

Second, with respect to any dimension of idealization, a theory can build in a more or less demanding idealizing assumptions. For example: consider informedness. We might idealize modestly, assuming that citizens are all (e.g.) informed about the constitutional basics of their polity. Or we might idealize more boldly, assuming that the polity is *free of misinformation*, or, more boldly still, that citizens are *omniscient*. Similar points apply to other dimensions of idealization. For example, it might be illuminating for some purposes to theorize about certain political questions under assumptions of perfect compliance, but, for other purposes, it might be illuminating to engage in political inquiry that assumes a non-actual but feasibly achievable level of compliance.[34] More generally, a theory can be more or less "realistic" or "utopian" in its assumptions about human motivation, the material conditions we face, etc.

Third, idealizations can be valuable for inquiry for different reasons. Consider three examples.

(a) An idealization might be *constitutive* of the normative category that we are investigating. For example, in *The Moral Problem,* Michael Smith argues that facts about what certain kinds of idealized agents (ones with more information, who do not make certain kinds of reasoning errors) would want our current "non-ideal" selves to do *constitute* facts about what normative reasons for action we have.[35] In political inquiry, certain idealizations might similarly be argued to be constitutive of a perfectly just regime.

(b) An idealization might be *epistemically valuable for inquiry*, by helping us to more efficiently arrive at relevant knowledge, understanding, or some

---

[34] In connection with this point, it is instructive to consider the rule consequentialism literature, where there is a vigorous debate concerning whether the relevant rules should be tested relative to (e.g.) *compliance* or *acceptance*, and what degree of compliance (etc.) is relevant. For discussion, see (Hooker, 2023, §6).

[35] (Smith, 1994).

other epistemic goal. For example, if we are wondering about how to evaluate the actions of a particular university's administration, it might be helpful to abstract away from certain aspects of the university's politics, in order to get a deeper critical perspective on the overall priorities of the university. Similarly, some have argued that understanding the "ideally just" is important for evaluating practical proposals for political improvement.[36]

(c) An idealization might be *explanatorily illuminating* in some further way. Suppose there are explanatory relations that do not simply reduce to epistemic or constitutive relations. If so, then idealizing facts could play one of these roles. (For example, on some metanormative views, certain naturalistic facts can *ground* normative facts, without partially *constituting* them.)[37] Idealizing facts might well play such a grounding role. For example, one might think that *what grounds* one (imperfectly just) state being preferable to another is their relation to certain facts about ideally just states.

Fourth, idealizations can sometimes be theoretically *unhelpful* or *counterproductive* in political inquiry, for several reasons. Consider three:

(a) An idealization might simply fail to help answer the question in political inquiry that most interests us. For example, Amartya Sen argues that, in order to evaluate real-world political outcomes and policies, we need to be able to evaluate the *comparative* justice of alternative outcomes we could achieve. He uses an analogy to argue that perfect justice is often simply irrelevant to this task: if I want to know which of two mountains is taller, knowing that some third mountain is the tallest in the world is orthogonal to my task. In a similar way, knowing that some state C is perfectly just may be totally unhelpful to knowing which of A or B is *more just*.[38]

---

[36] (Simmons, 2010, 34-36).
[37] For example, see (Berker, 2019).
[38] (Sen, 2006).

(b) Straightforward inference from normative premises about idealized circumstances to normative conclusions about realistic ones can be unreliable, in light of the "problem of the second best".[39] To adapt David Enoch's example, speed limits that would be optimal under full compliance with traffic laws might be disastrous given how people are actually disposed to drive.[40]

(c) Some idealizations eliminate the very features that partially constitute the most pressing political challenges that we face. This elimination, it has been argued, can play an ideological function.[41] As such, focusing our attention on debates about the norms that regulate such an ideal society can make it possible to engage systematically in normative political inquiry without ever addressing the pressing question of what justice requires *of us*, in the unjust here and now.

Together, the first two dimensions of variation suggest that there is an exceptionally broad range of possible degrees of idealization, with various target dimensions, that might appear relevant to political inquiry. The third and fourth dimensions of variation suggest that different nodes across this range will vary strikingly in whether and how they are valuable, useless, or counterproductive.

To bring this point out, consider two examples. First, *some* normative political questions are constitutively tied to idealizations, while others are not. Contrast, for example, the question of what is compatible with *perfect justice* with the question of whether a certain decision made by the Supreme Court in Canada is *more or less* just than an alternative decision it might have made. Even if Sen is correct that certain idealizations are irrelevant to answering the second question, they might be indispensable for answering the first question. Second, suppose that achieving meaningful reparations for slavery is in fact "politically infeasible" in the current social/historical circumstances in the USA.

---

[39] The foundational text for discussion about "the problem of the second best", in the context of theorizing pareto optimality, is (Lipsey & Lancaster, 1956-1957). For recent discussion about the importance of this problem for normative work in political philosophy, and for further references to other recent discussions, see (Wiens, 2020).

[40] See (Enoch, 2018).

[41] (Mills, 2005).

If true, this reality might imply that pushing for reparations is not a useful proximate goal for political action. But nonetheless, idealizing away from these feasibility facts might play a liberatory, as opposed to oppressive, role in our political thought.

4. The case against retention

In this section, we draw on the heterogeneous use data and the heterogeneous function data to argue against retention of the term 'ideal theory' and its cognates in political inquiry.

Before we begin, we want to note some general reasons in favor of retaining a piece of technical terminology, which apply to the current use of 'ideal theory'. These include benefits that flow from familiarity with the terminology and concepts (and what they refer to) within the relevant community of inquiry, and ones that flow from smoother engagement with the existing literature that employs that terminology and those concepts.

We want to emphasize that these benefits set a relatively low bar. For example, shortly after a more accurate theory of combustion was developed, use of 'phlogiston' retained these virtues. But, these virtues were clearly outweighed by other considerations, which favored abandoning use of the term 'phlogiston' for the purposes of scientific inquiry. These considerations included, crucially, that the new theory no longer posited the existence of phlogiston, and that people had no good independent evidence in favor of believing in phlogiston. We don't take the case for abandoning 'ideal theory' and its cognates for the purposes of political inquiry to be as strong as the case was for abandoning 'phlogiston' for the purposes of scientific inquiry. However, we argue that, as in the 'phlogiston' case, there are considerations that outweigh the modest considerations we just canvassed in favor of retention, and point in favor of abandonment.

Our argument draws on three central ideas: the standards for evaluating technical terminology, the unhelpfulness of the pattern of usage characterized in §2 above, and unreliable inferences that this existing pattern can lead to.

Our first line of argument appeals to the fact that 'ideal theory' and its cognates are *technical terms*, like 'supervenience' but unlike, say, 'justice'. We think this fact is relevant to how we evaluate whether to retain a piece of terminology in at least two ways.

The first issue is that there are dangers associated with abandoning folk terminology that do not apply to abandoning technical terminology. To illustrate this point, imagine for a moment a proposal that political inquirers abandon the use of 'justice'. There are several deep worries one might have about this proposal, in light of the fact that 'justice' is a central folk term for thinking about politics. For example, ceasing to use 'justice' would risk decoupling political inquiry from an important topic that ordinary people care deeply about with respect to social/political life.[42] Further, thinking in terms of 'justice' might be *epistemically crucial* for many people: it might be that much of our competence in thinking about normative political questions is psychologically encoded in terms of 'justice'. These worries do not seem to apply to abandoning a technical term.

The second way in which the fact that 'ideal theory' and its cognates are technical terms is relevant is that we characteristically introduce and retain technical terms in order to achieve certain broadly epistemic goals such as (1) the ability to track joint-carving distinctions and (2) increased clarity, precision, and usability.

Consider the first goal: sometimes it can be valuable to introduce and retain a technical term because doing so helps to carve the relevant domain at its joints, which we have epistemic reason to want to do.[43] Presumably, we successfully and usefully introduced 'atom' in this way, despite our having striking false beliefs about the nature of atoms for centuries after its

---

[42] For similar worries, see (Jackson, 1998), as well as (Strawson, 1963)'s famous concerns about Carnapian explications (of the sort discussed in (Carnap, 1947/1956)). For connected discussion, see (McPherson & Plunkett, 2021b).

[43] It should be noted that, on some ways of thinking about various possible values in conceptual ethics, a concept's being "joint-carving" is a primarily "metaphysical", rather than "epistemological", good. See, for example, the discussion in (Burgess & Plunkett, 2013b). We don't want to wade into debates here about how to best characterize these different goods. Our point here is just that, in many contexts, we might well have epistemic reasons to track an important metaphysical distinction.

introduction. One upshot of §3, however, is that there is no reason to believe that there is a single theoretical joint in the vicinity of contemporary uses of 'ideal theory'. Rather, what we see is a wide range of possible idealizations that could potentially be theoretically useful to track in particular contexts.

Now consider the second goal. Often, useful technical terms or explications have meanings that are *more precise and explicit* than alternative terminology, or are *more usable* (e.g. in having a clear operationalization that can be used for certain inquiries). But as we emphasized in §2, what we see in the case of 'ideal theory' and its cognates is instead a forest of cross-cutting distinctions. We conclude from this that contemporary use of 'ideal theory' lacks the sorts of virtues that typically support retaining a technical term.

The second central strand of our case for abandonment appeals to the pattern of contemporary usage summarized in §2. The simplest case for abandonment here is that this existing pattern is, frankly, a mess. But we can say more.

One possible consequence of the heterogeneity of usage is that the term 'ideal theory' and its cognates simply fail to refer. One way to motivate this idea is as follows: if the meaning of our words depends in some way on how relevant "experts" use words, and there is massive divergence in the expert usage of a given word, then that might yield the result that the word fails to refer.[44] If the term 'ideal theory' fails to refer, that would constitute a powerful and familiar reason for abandoning the term.[45]

Another credible possibility is that 'ideal theory' manages to refer, but is *polysemous*: that is, the term as ordinarily used has a range of different meanings (this idea is seemingly suggested by two of the survey articles we discussed in §2). Polysemy is plausibly a striking vice in a technical term, especially if (as is the case here) the relevant meanings are quite different from each other.

---

[44] For discussion of this sort of possibility, including an argument that it might apply to many more words than one might initially think, see (Cappelen, 2013).

[45] This interacts with the fact that 'ideal theory' and its cognates are technical terms. Some philosophers argue for retaining certain folk terms that they believe fail to refer because of the usefulness of other aspects of their widespread use. For an example of this, in the case of moral terminology, see (Olson, 2014).

Perhaps the most promising hypothesis for the retentionist is that, on current usage, 'ideal theory' and its cognates are *context-sensitive* expressions, with facts about the context of use fixing which of the extremely heterogeneous collection of apparent meanings for the term is expressed in that context. It is unclear how promising this is as a hypothesis about the actual semantics of 'ideal theory'. For our purposes, we will simply grant this hypothesis for the sake of argument and ask: if the hypothesis is true, does that give us reason to retain the term?

In our view, it is generally undesirable for technical terms to be context-sensitive, for several reasons. For example, a context-sensitive term will not consistently carve nature at the joints. And it will not provide precision and clarity in a transparent way. In light of these points, we think there has to be a specific rationale for embracing a context-sensitive technical term. It may be that it is sometimes fruitful to have technical terms that allow inquirers to coordinate on different intensions in different contexts.[46] But we see no such rationale in the context of 'ideal theory'. To foreshadow our constructive alternative: there is no evident practical need in political inquiry to coordinate in this way, as opposed to simply focusing on discussing and evaluating *specific* idealizations.

The final strand of our argument against retaining 'ideal theory' and its cognates begins with the idea that the heterogeneity of the existing pattern of usage associated with 'ideal theory' may promote various sorts of unreliable inferences.[47] The basic concern here is as follows. The heterogeneous use data suggest that a wide variety of contrasting uses of 'ideal theory' are salient in contemporary political inquiry. This variation raises a danger that inquirers could make unreliable inferences, by drawing conclusions based upon unintended but psychologically salient uses.

---

[46] For an argument for this kind of view about the term 'topic continuity' in conceptual engineering (and beyond), see (McPherson & Plunkett, 2021b).

[47] The worry we describe below about unreliable inferences connected to uses of 'ideal' and 'non-ideal' theory is an instance of the kinds of general worries about unreliable inference that we discuss in (McPherson & Plunkett, 2020).

Consider one potential example of this sort of unreliable inference. When making idealizations, a normative theorist should be careful that these idealizations – and the rationales for deploying them – fit together in a coherent way. For example, one shouldn't be assuming some things that are wildly infeasible, while insisting on high standards of feasibility for other dimensions – at least without a good explanation for this discrepancy.[48] Framing one's discussion in terms of "ideal theory", rather than the particular idealizations at issue, can prevent us from carefully tracking how well the idealizing assumptions one is deploying fit together.

Consider a different way that 'ideal theory' may promote unreliable inference. The various heterogeneities discussed in §2 and §3 strongly suggest that there is not a single, theoretically interesting topic in the vicinity of existing discussions of "ideal theory". But the existence of a technical term can encourage the illusion that there is such a topic. One upshot of this illusion is that various authors sometimes disagree about whether certain topics are part of "the" "ideal theory/non-ideal theory" debate. In §2, we mentioned two examples of this: disagreement over the classification of Cohen's fact-insensitivity thesis, and disagreement over the classification of theories that are sensitive to issues of feasibility. The problem is this. We have suggested that there is no underlying theoretical unity in existing talk of "ideal theory". But in the absence of such unity, it is unclear what compelling theoretical basis there could be for insisting that there is some mistake being made in either including or excluding these topics from the topic of "ideal theory".

One further reason that unreliable inference is a substantial danger is that the considerations that might justify one kind of idealization (say, assuming conditions of full compliance in the Rawlsian vein) given one kind of project (say, trying to get at general normative principles that govern fair cooperation in a pluralistic society) won't always translate over to *other* idealizations for *other* projects. If we think of all of the projects as instances of

---

[48] For a case study involving how (often subtle) switches in the kinds of idealization used in making arguments in normative political philosophy can matter to assessing those arguments, see (Plunkett, 2021), discussing (Rose, 2020).

"ideal theory", we run the risk of thinking that they must all share a rationale, and thereby ignoring the justificatory structure relevant to our specific project.

Another kind of danger arises *among* inquirers. As we have seen, political inquirers frequently use the term 'ideal theory' to pick out *different* idealizing assumptions. Given this context, there is a clear worry that framing discourse in terms of "ideal theory" (or "non-ideal" theory) may obscure the different idealizing assumptions that different speakers are (or are not) making in a particular context. This lack of clarity in turn may render communication using the terms less fruitful.

We can pull all of these strands together to summarize our case against retaining 'ideal theory' as follows: 'ideal theory' and its cognates are technical terms, but lack the virtues that characteristically warrant the use of such terms. The pattern of usage for these terms may suggest reference failure or polysemy. The best-case interpretation – that the term is context-sensitive – is an unattractive feature for a technical term in this context. And the actual patterns of use raise dangers of unreliable inferences and confusion that we should seek to avoid with our technical terminology. We take these considerations to constitute a strong prima facie case against retention.

5. The case against reform

In this section, we argue against the idea that we should seek to *reform* existing usage of 'ideal theory' and its cognates rather than *abandon* use of the terms. Many inquirers discussing "ideal theory" and "non-ideal theory" state how they want their usage of 'ideal theory' and its cognates to be read in the context of their work. These suggestions might well be used as the basis for a potential broader reform of the terminology beyond the context of just their work. We cannot address every possible reforming proposal here. Instead, we first offer general reasons for pessimism about useful reforms, and then explain our pessimism about what we take to be two of the most attractive reform proposals.

We take the functional heterogeneity data canvassed in §3 to cast doubt on the general possibility of an attractive reform. The core reason is as follows. We think it is plausible that *certain* idealizations can sometimes be epistemically useful for each of the three reasons discussed in §3, in some contexts in political inquiry. But we think that it is highly implausible that there is a *specific collection* of idealizations that is useful across a *wide range* of topics in political inquiry, in a *wide range* of contexts. For example, we think that it is implausible that a specific collection of idealizations (each to a particular degree) is *constitutive* of any wide range of important normative categories of interest to political inquiry.[49] We think similar skepticism is warranted about there being a unified pattern of idealizations that are explanatorily or epistemically useful.

If this is so, then any specific way of reforming the content of 'ideal theory' looks like it will likely produce either a gerrymandered mess (if it adequately captures the range of central roles for idealizations), or a theoretically dubious ad hoc cut (if it does not). These issues suggest at first blush that reformed 'ideal/non-ideal' language would not be epistemically useful to inquiry in political inquiry.

The same points hold if we focus on the three ways that idealizations can be unhelpful or pernicious (discussed in §3). Again, it is very plausible that *certain* idealizations can sometimes be theoretically unhelpful or pernicious in each of these three ways. But the diversity of idealizations and how they interact renders it implausible that there is a non-gerrymandered way of reforming 'ideal theory' such that this term can express an "alertive" concept that enables us to reliably avoid these dangers. We could, of course, try to reform 'ideal theory' so that it refers to all and only the theories that involve pernicious idealizations. But, again, this approach would likely be gerrymandered and unmotivated.

This discussion suggests that there isn't a distinctively theoretically useful cut in the vicinity of the usage of 'ideal theory' (or: 'non-ideal theory') for

---

[49] This is a natural place for certain philosophers (e.g., orthodox Rawlsians) to resist our argument, given other commitments they have.

us to regiment to. We think that this result is unsurprising, given the heterogeneous use data: if there really were a single theoretically important cut, we would expect competent inquirers to eventually converge upon it. Instead, what we see in the recent literature is, if anything, an accelerating proliferation of cross-cutting usages.

We can complement this argument with a practical argument about the asymmetry between reform and abandonment. Suppose an individual chooses to abandon the use of a term. She might well want to recruit others to the cause – perhaps the more the better. But even if she does not get others to go along with her, she might still reap many of the benefits of abandonment in her own conversations: for example, both she and her interlocutors might avoid the kinds of "unreliable inference" worries we glossed above.

Contrast this situation with seeking to reform the use of a term. An inquirer might want to only reform the use (or meaning) of a term as it is used in a given context: for example, its use in their own article. (Perhaps the urge to do this is part of what explains the proliferation of competing characterizations of 'ideal theory' and its cognates.) Even that arguably takes *somewhat* more work to pull off than abandonment, and comes with more risks (e.g., continued vulnerability to the sorts of "unreliable inferences" mentioned in the previous section).

Moreover, many who seek reform want to reform the meaning (or use) of a term more generally, beyond just their own usage. That requires getting a critical mass of people to go along with the reform. To significantly reform the use of 'ideal theory' in this more ambitious way would, at the very least, require the investment of significant resources. In some cases, this sort of investment might be worthwhile. For example, if one thinks that a certain reform to existing patterns of use of race and gender terms would have non-trivial liberatory payoffs, those potential payoffs might warrant considerable investment into bringing about that reform.[50] But in the case of 'ideal theory'

---

[50] We take this to be a reasonable reading of a key part of the rationale for Haslanger's work in (Haslanger, 2000).

and its cognates, to put it bluntly, why would we bother doing this, if (as we have been arguing) the theoretical payoffs look so meager?

We now complement these general points with a discussion of two specific possible reforming proposals. The first is to simply reform usage so that it conforms with the introductory stipulations offered by Rawls. There are three reasonable motives for focusing on this reform. First, the relevant usage is familiar to participants in the discourse: almost everyone in contemporary political inquiry knows about Rawls' stipulations. Second, given Rawls' historical influence, one might hope that this reform would preserve continuity with a large fragment of the extant literature. Finally, on some views about language, one might insist that Rawls' successful introduction attempt *fixed the meaning* of the terms, and so this sort of reform to our usage would (desirably) lead participants to stop using these terms *incorrectly*.

Our core objection to this proposal is that the discussion in §3 suggests reasons to doubt that there is anything especially useful about Rawls's terminology. That is, it is far from clear that using the term 'ideal theory' to group together theories that assume full compliance, and circumstances hospitable to the realization of a certain conception of justice, carves the space of political views at its joints. When we combine this uncertainty with our general point above about the costs of attempting reform, we find it implausible that it is worth investing resources in trying to turn the clock back to have the meaning of 'ideal theory' and its cognates line up precisely with Rawls's original usage. And this is especially true because – without reform – one can make it perfectly clear that one wants to talk about "ideal theory in the sense spelled out in *A Theory of Justice*". As this makes clear: abandoning the terminology does not make the Rawlsian *idea* (or any of the other ideas that have traveled under this banner) unavailable to theorists.

A second proposal is to reform to something broader than Rawls's specific proposal, but which – one might think – captures the *genera* of which Rawls' theory is a species. To illustrate, consider what we take to be one of the most promising versions of this idea: to regiment to the idea that an "ideal theory" of x is a theory of something that is *perfectly x*. To see the basis for this

23

idea, consider Rawls' gloss on ideal theory as providing an account of a "perfectly just" basic structure of society.[51] The proposal on the table involves a genera notion, because different inquirers might disagree about (e.g.) what sorts of idealizations are constitutive of perfect justice. On this reform, such theorists would be understood as engaging in substantive disputes within "ideal theory".

We have three worries about this sort of proposal. First, it is not clear that this proposal has the sort of theoretical virtues that would warrant efforts to reform existing usage. Second, it is not at all clear that inquirers who build different idealizations into their notion of "perfect justice" are in fact engaged in substantive disputes about what perfection in justice consists in. For someone might be right that x is just relative to idealization A, and someone else might be right relative to idealization B. What is not clear is that there is anything of substance to the idea that idealization A – and not idealization B – is constitutive of *perfection* in justice. Third, recall an earlier point we made: on many contemporary usages, "non-ideal" theory, is not just any kind of theory that isn't "ideal theory". Rather, theorists posit a variety of competing specific relations between "ideal" and "non-ideal" theory. So, to be complete, this proposed regimentation of 'ideal theory' and its cognates would need to take a stand on which of these ways to regiment use of the term 'non-ideal'. But again, it is not clear that there is any distinctive theoretical payoff for such regimentation, which would warrant the effort needed to successfully reform usage.

6. Conclusion

In this paper, we've given an initial case for abandoning using the term 'ideal theory' and its cognates in the context of political inquiry. We can briefly sum up the case in this way. Start with the fact that the existing use of these terms is characterized by an alarming range of cross-cutting ideas. Further, we see a

---

[51] (Rawls, 1971/1999, 216).

broad range of idealizations that might be useful to political inquirers, and a diverse range of virtues and vices that such idealizations can exhibit across different contexts. On this basis, we have argued against A) *retaining* the existing pattern of use (because it exhibits significant vices, and lacks the characteristic virtues of technical terminology), and also B) seeking to *reform it* (because there is nothing theoretically important to associate with these words that warrants the necessary investment of resources to get inquirers to converge on a regimented pattern of use).

The conclusion of our main argument raises a natural question: if we are to cease using "(non-) ideal theory" talk, how should we be talking about the relevant issues instead?

In our view, there is an attractive way forward after abandonment. When engaging in political inquiry, we should focus our attention on the *particular* idealizations that are being made in a given argument, theory, etc. – or, somewhat more broadly, the particular *kinds* of idealizations. As part of this endeavor, we can focus on which dimensions of the context under consideration are being idealized – e.g., whether the idealizations concern the compliance of agents with institutions and laws, the emotions of agents, the availability of natural resources, etc. We can also make explicit *how* much idealization we are assuming on a given dimension. With this specificity of focus, it will then be natural to speak directly to *why* certain idealizations (and not others) are being made. And it will also (often) be natural to explain why those idealizations are distinctively relevant to our inquiry, why they are not vulnerable to the problem of the second best, and why they are not ideologically pernicious. There would be no illusion that, by using the term 'ideal theory' (or 'non-ideal theory'), we were somehow in a position to assume that these questions had already been answered.

Doing these things, we claim, holds the promise of avoiding all of the main issues we introduced above for contemporary discussion of "ideal" and "non-ideal" theory, or at least doing better on this front than we could do by continuing to use this terminology. Further, given the heterogeneity of the existing literature, and that of the theoretical function of idealizations, we don't

see what of value would be lost on our alternative approach besides a certain kind of discontinuity with the existing literature. Especially given how heterogeneous the use of 'ideal theory' and its cognates is in the existing literature, it's not clear if that is a meaningful loss at all.

In closing, it is worth (re-)emphasizing and developing a point we made in the Introduction. Our argument for abandoning "ideal theory" talk is orthogonal to the evaluation of the practice of idealization in political inquiry more generally. To take the case of the central term 'ideal theory': we are arguing for abandoning the use of that term, not against the use of *idealization* in political inquiry, or against doing the kind of inquiry that some characterize as "ideal theory". We think that many of the debates that currently travel under the banner of "ideal vs. non-ideal theory" involve issues of substantial theoretical importance. We also think that theories routinely characterized using 'ideal theory' and its cognates can be worth investigating and developing. We just think that it would be better to talk about these topics, and pursue these projects, with terminology that does not suffer the vices we have identified for 'ideal theory' and its cognates.

Let us illustrate this point. We are sympathetic to many of the dangers of idealization sketched in §3. These dangers are often used as the basis for criticisms of the project of "ideal theory". And an important recent trend in political inquiry involves quick dismissals of all "ideal theory" as *unnecessary* for normative guidance, *irrelevant* to the evaluation of our actual circumstances, objectionably *ideological*, etc.[52] The points we have made in this paper suggest reasons for skepticism about these conclusions, when they are understood to target all political inquiry that makes use of idealizations.

We grant that each of the dangers of idealization has merit when applied to certain specific patterns of idealization. But the functional heterogeneity data suggests that we should be skeptical that these dangers apply to the use of idealization in political inquiry quite generally. In light of this, we take blanket dismissal of "ideal theory" to be another example of a kind

---

[52] See (Mills, 2005) for an important statement of these kinds of concerns about "ideal theory".

of unreliable inference, invited by the illusion of unity produced by organizing inquiry around a theoretically unhelpful technical term.

Obviously, we haven't given anything like a conclusive argument for abandoning 'ideal theory' and its cognates in this short paper. But we hope that our argument is powerful enough to give pause. If it succeeds in prompting readers to think more about why we are giving this terminology a central place in our thought and talk about politics, we'll take that as a win.

Acknowledgements

WORKS CITED

Anderson, E. (2010). *The Imperative of Integration*. Princeton University Press.

Berker, S. (2019). The Explanatory Ambitions of Moral Principles. *Nous*, *53*(4), 904-936.

Burgess, A., Cappelen, H., & Plunkett, D. (2020). *Conceptual Engineering and Conceptual Ethics*. Oxford University Press.

Burgess, A., & Plunkett, D. (2013a). Conceptual Ethics I. *Philosophy Compass*, *8*(12), 1091-1101.

Burgess, A., & Plunkett, D. (2013b). Conceptual Ethics II. *Philosophy Compass*, *8*(12), 1102-1110.

Burgess, A., & Plunkett, D. (2020). On the relation between conceptual engineering and conceptual ethics. *Ratio*.

Cappelen, H. (2013). Nonsense and Illusions of Thought. *Philosophical Perspectives*, *27*(1), 22-50.

Cappelen, H. (2023). *The Concept of Democracy: An Essay on Conceptual Amelioration and Abandonment*. Oxford University Press.

Cappelen, H., & Dever, J. (2021). On the Uselessness of the Distinction between Ideal and Non-Ideal Theory (at least in the Philosophy of Language). In J. Khoo & R. K. Sterken (Eds.), *Routledge Companion to Social and Political Philosophy of Language* (pp. 91-106). Routledge.

Cappelen, H., & Plunkett, D. (2020). A Guided Tour of Conceptual Engineering and Conceptual Ethics. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.),

Conceptual Engineering and Conceptual Ethics (pp. 1-26). Oxford University Press.

Carnap, R. (1947/1956). *Meaning and Necessity: A Study in Semantics and Modal Logic*. University of Chicago Press.

Chalmers, D. (2020). What is Conceptual Engineering and What Should it Be? *Inquiry*.

Cohen, G. A. (2003). Facts and Principles. *Philosophy and Public Affairs*, *31*(3), 211-245.

Cohen, G. A. (2008). *Rescuing Justice and Equality*. Harvard University Press.

Enoch, D. (2018). Against Utopianism: Noncompliance and Multiple Agents. *Philosophers' Imprint*, *18*.

Geuss, R. (1981). *The Idea of a Critical Theory : Habermas and the Frankfurt School*. Cambridge University Press.

Haslanger, S. (2000). Gender and Race: (What) Are They? (What) Do We Want Them to Be? *Nous*, *34*(1), 31-55.

Hooker, B. (2023). Rule Consequentialism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy (Spring 2023 Edition)*. https://plato.stanford.edu/archives/spr2023/entries/consequentialism-rule/.

Isaac, M. G. (2020). How To Conceptually Engineer Conceptual Engineering? *Inquiry*, 1–24.

Jackson, F. (1998). *From Metaphysics to Ethics: a Defence of Conceptual Analysis*. Clarendon.

Jaggar, A. M., & Tobin, T. W. (2013). Situating Moral Justification: Rethinking the Mission of Moral Epistemology. *Metaphilosophy*, *44*(4), 383-408.

Lipsey, R. G., & Lancaster, K. (1956-1957). The General Theory of Second Best. *The Review of Economic Studies*, *24*(11), 11-31.

McKean, B. (2017). Ideal Theory after Auschwitz? The Practical Uses and Ideological Abuses of Political Theory as Reconciliation. *Journal of Politics*, *79*(4), 1177-1190.

McPherson, T., & Plunkett, D. (2017). The Nature and Explanatory Ambitions of Metaethics. In T. McPherson & D. Plunkett (Eds.), *The Routledge Handbook of Metaethics* (pp. 1-25). Routledge.

McPherson, T., & Plunkett, D. (2020). Conceptual Ethics and The Methodology of Normative Inquiry. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.), *Conceptual Engineering and Conceptual Ethics* (pp. 274-303). Oxford University Press.

McPherson, T., & Plunkett, D. (2021a). Conceptual ethics, metaepistemology, and normative epistemology. *Inquiry*.

McPherson, T., & Plunkett, D. (2021b). Topic Continuity in Conceptual Engineering and Beyond. *Inquiry*.

McPherson, T., & Plunkett, D. (2022). Normative standards and the epistemology of conceptual ethics. *Inquiry*.

Mills, C. W. (2005). "Ideal Theory" as Ideology. *Hypatia*, *20*(3), 165-184.

O'Neill, O. (1996). *Towards Justice and Virtue: A Constructive Account of Practical Reasoning*. Cambridge University Press.

Olson, J. (2014). *Moral Error Theory: History, Critique, Defence*. Oxford University Press.

Pinder, M. (2021). Conceptual Engineering, Metasemantic Externalism and Speaker-Meaning. *Mind*, *130*(517), 141–163.

Plunkett, D. (2016). Conceptual History, Conceptual Ethics, and the Aims of Inquiry: A Framework for Thinking about the Relevance of the History/Genealogy of Concepts to Normative Inquiry. *Ergo*, *3*(2), 27-64.

Plunkett, D. (2021). The Varieties of Idealization and the Politics of Economic Growth: A Case Study on Modality and the Methodology of Political Philosophy. *Inquiry*.

Rawls, J. (1971/1999). *A Theory of Justice (revised edition)*. Harvard University Press.

Rose, J. L. (2020). On the value of economic growth. *Politics, Philosophy and Economics*, *19*(2), 128-153.

Scharp, K. (2013). *Replacing Truth*. Oxford University Press.

Sen, A. (2006). What Do We Want from a Theory of Justice? *Journal of Philosophy*, *103*(5), 215-238.

Shelby, T. (2016). *Dark Ghettos: Injustice, Dissent, and Reform*. Harvard University Press.

Simmons, A. J. (2010). Ideal and nonideal theory. *Philosophy and Public Affairs*, *38*(1), 5-36.

Smith, M. (1994). *The Moral Problem*. Blackwell.

Stemplowska, Z. (2008). What's Ideal About Ideal Theory? *Social Theory and Practice*, *34*(3), 319-340.

Stemplowska, Z. (2017). Non-Ideal Theory. In K. B. Kasper Lippert-Rasmussen, and David Coady. (Ed.), *A Companion to Applied Philosophy* (pp. 284-296). John Wiley & Sons.

Stemplowska, Z., & Swift, A. (2012). Ideal and Non-Ideal Theory. In D. Estlund (Ed.), *The Oxford Handbook of Political Philosophy* (pp. 373–388.). Oxford University Press.

Strawson, P. F. (1963). Carnap's Views on Conceptual Systems versus Natural Languages in Analytic Philosophy. In P. A. Schilpp (Ed.), *The Philosophy of Rudolf Carnap* (pp. 503--518). Open Court: La Salle.

Tessman, L. (2009). *Feminist Ethics and Social and Political Philosophy: Theorizing the Non-Ideal*. Springer.

Thomson, C. (2020). Ideal and Nonideal Theory in Political Philosophy. In *Oxford Research Encyclopedia of Politics*.

Valentini, L. (2012). Ideal vs. Non-ideal Theory: A Conceptual Map. *Philosophy Compass*, *7*(9), 654-664.

Wiens, D. (2020). The General Theory of Second Best Is More General Than You Think. *Philosophers' Imprint*, *20*(5), 1-26.