Res Publica (2024) 30:189–210
https://doi.org/10.1007/s11158-023-09597-0

Is Approximation of an Ideal Defensible?

Dai Oba[1]

Accepted: 17 April 2023 / Published online: 14 June 2023

Abstract

What role does our knowledge about the ideal society play in guiding policymaking in the real world? One intuitive answer is to approximate. Namely, we have a duty to approximate the ideal within the relevant constraints of feasibility. However, political philosophers seem to have what might be called 'approximatophobia'. Many philosophers, including idealists such as David Estlund, warn against approximation. Their criticism is chiefly motivated by 'the problem of second best', which points out that your second-best option may not be closest to your best option. This paper aims to dispel 'approximatophobia'. The difficulty posed by the problem of second best is often overstated. More positively, I present a novel defence of approximation, arguing that approximation of an ideal can be a reasonably reliable default strategy of action guidance in the real world. Difficulties that may afflict the project of approximation can be mitigated by sophistication of the project of approximation. After showing that critics of approximation overstep their mark in issuing the strong or moderate warning against approximation, I propose an account of sophisticated approximation. It seeks a series of reforms that make existing social institutions closer to the ideal based on careful selection of frame of analysis. Three virtuous correlations are identified as indicators for successful project of approximation, linking descriptive similarity to desirability, feasibility and knowledge of the ideal. I also explain a two-staged strategy of sophisticating the way you approximate the ideal, with an expected positive feedback effect.

Keywords Non-ideal theory · Action guidance · Justice · Feasibility · Second best

✉ Dai Oba
dai.oba@politics.ox.ac.uk

[1] Department of Politics and International Relations, St Catherine's College, University of Oxford, Manor Road, Oxford OX1 3UJ, UK

Introduction

In our pursuit of social justice, ideal theories offer guidance that seems at least initially plausible (Wolff 2015). Of course, ideal theories have been criticised for ignoring feasibility constraints and other important features of the real world (e.g., Geuss 2016; Mills 2017).[1] So, what we should aim for is the best approximation of the ideal. We—people in the real world—have an interest of justice to approximate the ideal within the relevant constraints of feasibility. Or so you would think.

In fact, the approximation of an ideal is quite unpopular among political philosophers, including those who defend ideal theories. Amartya Sen, Robert Goodin, David Wiens, Jonathan Wolff and David Estlund all warn against approximation. At the core of their criticism is the problem (or theory) of second best,[2] which says that the second-best option may not be the one most descriptively similar to the best option. Sen's famous example is wine. People who prefer red wine to white wine do not prefer the mix of the two to white wine, although it is descriptively more similar to red wine (Sen 2009, p. 16). In the same vein, when we approximate our social institutions to the ideal, we are not necessarily making them more desirable (Goodin 1995, pp. 53–54). Short of the full realisation of the ideal, aiming for a more complete realisation of the ideal institutional arrangement is a misleading guide to improving social institutions. Should we abandon the approximation of an ideal? I say no.

The aim of this paper is to dispel the exaggerated doubt cast on the project of approximation (you might call it 'approximatophobia'). The magnitude of the difficulty posed by the problem of second best should not be overstated. More positively, I propose and defend sophisticated approximation of an ideal. We can often distinguish well-informed, careful approximation from crude approximation, which all too obviously fails. I argue that carefully executed approximation of an ideal has pro tanto validity as the default strategy of action guidance in the real world. Additionally, we may expect an approximation of an ideal to make the ideal more feasible and take on a heuristic function, which is important since our knowledge about social phenomena is fundamentally limited.

The structure of this paper is as follows. The first section ('Examining Anti-approximation Arguments') examines arguments against approximation. After clarifying different degrees of caution against approximation, I argue that the existing critique of approximation offers limited support to substantive cautions. The second section ('Sophisticated Approximation') presents a positive argument and develops an account of the sophisticated approximation. I will show how sophisticated approximation of an ideal can minimise the difficulties posed by critics of

[1] For assessment of criticisms to ideal theory itself, see Sirsch (2020). Also, Volacu (2018), Arvan (2019) and Sirsch (2020) provide a larger framework of using ideal theory in non-ideal context, to all of which this paper can be positioned as offering a more specific strategy of guidance to policymaking.

[2] It was first put forward by Lipsey and Lancaster (1956) in the context of the market. However, this paper focuses on the discussions of this problem by political philosophers, which may not necessarily follow economists' formulations (Goodin 2012; Wiens 2016b).

approximation. The careful selection of a frame of analysis is crucial to the sophisticated approximation. I will also identify signs of good approximation, focusing on three variables of desirability, feasibility and knowledge. The third section ('Objection and Alternatives') considers alternative strategies of action guidance: robustness strategy, case-by-case empiricism and countervailing deviation. I argue that none of these alternatives outperforms approximation as a default strategy of action guidance in the real world.

Let me clarify the focus of the paper.

First, this paper follows Rawls's terminologies of the ideal and non-ideal theory. An ideal theory attempts to identify principles of justice for an idealised society whose conditions are reasonably favourable and whose citizens are generally compliant with the principles of justice.[3] But the ideal theory still takes as given general facts about society as well as about human abilities and psychology. A non-ideal theory considers the requirements of justice in a society where the ideal is not fully realisable due to various constraints.[4] While I understand ideal theory to include insights about relations of values, this paper is not concerned with Cohenite ideal theories or what Hamlin and Stemplowska (2012) call the 'theory of ideals'.

Though there are various different types of agents or actions (Swift 2008; Stemplowska 2016), I focus on the collective actions of a society in the forms of public policies and institutions. Accordingly, this paper brackets considerations of personal costs and non-justice interests of individual agents.[5] Similarly, this paper sets aside the issue of possible disagreement about what the ideal is. For the sake of simplicity, I assume Rawls's well-ordered society and property-owning democracy (or liberal socialism) to be the relevant ideal.[6]

Next, by approximation of an ideal, I mean the partial realisation of the ideal that is nonetheless the most perfect within the feasibility constraints. Regarding social institutions, approximation of the ideal means choosing from the available institutional arrangements the one most descriptively similar to the ideal institutional arrangement in the relevant aspects (I will later discuss what the relevant aspects might be). Thus, in my terminology, approximation is a greater realisation of the descriptive features of policies and institutions, including rules governing them rather than the realisation of some abstract value (e.g., justice) itself. I argue that approximation in this sense can be effectively action-guiding in the real, non-ideal world.

[3] Marginal non-compliance such as conscientious objections and civil disobedience is consistent with the ideal society (Rawls 1999, pp 320–323).

[4] More specifically, Rawls distinguishes partial compliance of the principles of justice and unfavourable conditions as characteristics of the non-ideal society (Rawls 1999, p. 216). Some scholars make finer distinctions about types of feasibility constraints (e.g., Gilabert 2017; Sirsch 2020; Estlund 2011; Wiens 2016a; Hamlin 2017; Guillery 2021). For the present purpose, the rough distinction of the ideal and non-ideal should suffice, since my argument in this paper does not hinge on detailed specification of what the ideal (theory) is. For example, while I understand ideal theory along the lines of Simmons (2010), my argument is fully compatible with revisions to ideal theories suggested in Mills (2017).

[5] I will discuss the costs of identifying the right course of policy/institutional reform.

[6] While I think my framework can accommodate different ideals and some disagreement about ideals, this paper does not aim to demonstrate this.

Finally, the condition of real-world policymaking is such that time, knowledge and other resources are limited, and information about the desirability or full consequences of available options is incomplete. Default strategy has the pragmatic virtue of being attuned to such conditions. It determines what should be done *prima facie* while leaving room for adjustment. It specifies the first option to be considered or a course of action to be followed unless you have credible evidence against it. Action guidance[7] is rarely discussed this way, but given that no strategy works all the time, we should ask: what strategy is presumptively reliable?

Examining Anti-Approximation Arguments

This section discusses a prominent form of anti-approximation argument that relies on the problem of second best. I argue that there is limited support for anti-approximation warning even in a moderately strong form, calling into question much of its substantive force to disqualify approximation as a default strategy of action guidance in the real world. My aim here is to suggest that the jury is still out on the reliability of approximation.

Anti-Approximation Warnings of Different Magnitudes

In general terms, the problem of second best says that similarity is not a reliable sign of desirability. While critics of approximation commonly appeal to the problem, its lesson to approximation remains somewhat ambiguous. Let us start by disambiguating different magnitudes of warnings against approximation:

> Very strong warning: An approximation of an ideal *never* brings about the optimal[8] situation.
> Strong warning: An approximation of an ideal is *likely to fail* to bring about the optimal situation.
> Moderate warning: An approximation of an ideal *occasionally fails* to bring about the optimal situation.
> Weak warning: An approximation of an ideal does *not necessarily* bring about the optimal situation.

The stronger the warning, the harder it is to defend and the more substantive its implications. On one hand, while the weak warning is the easiest to defend, it does not mean much more than the logical possibility of failure: we cannot conclusively exclude the possibility that approximation may fail. In the same sense, we cannot exclude the possibility of aliens invading the Earth, but in the domain of ordinary public policy, it does not merit our attention. On the other hand, the strong warning seems sufficient to disqualify approximation as a guide to policymaking. It is simply

[7] I loosely follow North's notion of action guidance, whose major requirements involve coherence, reliability, determinateness and wide applicability (2017, pp. 78–79).
[8] I use 'optimal' to mean the second best here.

unwise to rely on a strategy that is unlikely to work. While less conclusive, the moderate warning certainly discredits the reliability of approximation.

Philosophers generally favour the weak and moderate warnings, but they seem undecided between the two. Goodin is a case in point. He clearly leans towards the weak warning, emphasising that the point of the problem of second best is the lack of a guarantee that the optimal feasible option resembles the ideal option (Goodin 2012, pp. 157–172). However, he also suggests that the threat is more than a logical possibility. He says:

> The trouble that the Theory of Second-Best makes for policy choice is not easily avoided. It issues firm cautions against the strategy of bringing the real world more into line with the presuppositions of ideal theory […] (Goodin 2012, p. 172)[9]

Wording such as 'not easily avoided' and 'firm cautions' suggests that Goodin is issuing the moderate (if not the strong) warning. So, he seems to shift between the weak warning and moderate warning without clearly settling on one or the other.

Wiens and Estlund sing similar tunes. While they mostly confine themselves to the weak interpretation, they also suggest that the problem of second best is hard to avoid or contain. Wiens says that 'the theory of second best poses a stiff *anti-approximation warning*' (Wiens 2016b, p. 133, original emphasis). Estlund also claims 'that it cannot be assumed that it is, or even probably is, an appropriate practical goal to approximate' a social ideal (Estlund 2020, p. 272). Despite Estlund's cautious phrasing, the suggestion that the approximation of an ideal is inappropriate as a practical social goal goes beyond the weak warning. Or, perhaps Estlund does not even intend to suggest this. For, logically speaking, denying that approximation 'probably is' appropriate as a policy aim does *not* of course entail that approximation is (probably) *inappropriate* as a policy aim.[10] That is, Estlund might be committed only to the weak warning. But then, something does not seem to add up, because in that case he would have to believe that the problem of second best is a mere logical possibility, like the invasion from space. And it 'probably is' appropriate to assume that an alien invasion does not concern our public policy. It seems Estlund still has to account for the difference between an alien invasion and the problem of second best, and thus needs something like the moderate warning.

In any case, at least for Goodin and Wiens, the ways they talk about the problem of second best suggests that they, perhaps implicitly, endorse the moderate warning. If they did not subscribe to something more than the weak warning, they would seem to lack a sufficient reason for their 'firm' or 'stiff' anti-approximation warning. At the very least, defending the moderate warning is the most straightforward avenue for the critics of approximation.[11]

[9] It is worth noting that Goodin makes this statement, being aware of the possible variations in the strengths of the anti-approximation warnings like I laid out above (Goodin 2012).

[10] This reading suggests also the possibility that Estlund's position is consistent with adopting approximation as a default strategy. I will come back to this in the penultimate section.

[11] Alternatively, they could take an anti-approximation view based on the weak warning combined with a deep empiricist suspicion of presumptions, which is considered in the penultimate section.

Moderate Warning: Conjectural Explanations and Examples

So, the critics of approximation vacillate between the highly plausible, weak warning and something more substantive, namely the moderate warning. However, no principled argument for the moderate warning seems to exist. Instead, philosophers offer conjectures about how the problem of second best might occur, using some illustrative examples. Do they do any good? Maybe. If there are sufficiently compelling explanations and examples about situations where the problem of second best occurs, they could attest—albeit speculatively—to the validity of the moderate warning.

Before examining this possibility, we need to be a bit more precise about the moderate warning. It states that the approximation of an ideal *occasionally fails* to realise the optimal situation. First, the word 'occasionally' indicates a non-negligible probability; unlike an alien invasion, it does sometimes happen. As this statement involves an empirical judgement, the quality and quantity of examples may matter in deciding whether the moderate warning can be sustained. Second, the word 'fail' must mean something along the lines of 'to encounter a difficulty that is not easy to avoid or solve'. If the difficulty can be avoided or solved, the problem of second best does not constitute a sufficient threat to approximation. To summarise these observations, for a compelling case to be made for the moderate warning, examples need to satisfy two conditions: they need to show that the problem of second best occurs sufficiently frequently and that these occurrences are hard to avoid. Call these factors the *prevalence condition* and the *inescapability condition*.

There are three possible explanations for how the problem of second best might realistically occur, along with examples to support them: suboptimisation, interdimensional interactions and incomplete ranking.

The first explanation is what Goodin calls suboptimisation. You make the error of suboptimisation when you try to approximate an ideal only along some but not all relevant aspects (Goodin 2012, pp. 157–159). Goodin offers an example. You need to select a car where three variables are relevant: (1) production date, (2) colour and (3) brand. Your ideal vehicle is a (1) new (2) silver (3) Rolls-Royce, but sadly, it is not available. Among available options, you rank a (1) week-old (2) black (3) Jaguar over a (1) new (2) silver (3) Toyota because you want a luxury car. However, the strategy of approximation will recommend the Toyota rather than the Jaguar. Hence, approximation misleads. Goodin puts the lesson succinctly: 'Typically, the right thing to do all-things-considered differs from the right thing to do only-some-things-considered' (Goodin 2012, p. 159).

Regarding the inescapability condition, there are some solutions that can mitigate the difficulty. In Goodin's example, you know that you want a luxury car. You can naturally infer from this that brand is a weighty criterion. Thus, you should prioritise it. In other words, you should approximate the ideal in a way that gives due weight to important aspects. The example can be made more complex. Perhaps you are unaware of how important the dimension of brand is or even that it matters at all, although you are subconsciously obsessed with luxurious brands. However, in such a case, there is a simple—theoretically simple, that is—solution. You need a better understanding of the ideal so that you know the major reason(s) why you value the

ideal. Therefore, the solutions would be to first have a good understanding of the ideal and then apply that knowledge to prioritise relevant aspects for approximation. These solutions would alleviate, if not eradicate, the problem of suboptimisation.

The same applies to another influential example, i.e., Sen's wine mixing case. Notice that no connoisseur would think that mixing the red and the white might be a good idea. If anything, as the practice of blind-tasting suggets, colour may be the first thing to ignore for accurate perception of taste. Thus, not only does this example not count in favour of prevalence of difficult problems of second best, it suggests that problems of second best may be easy to avoid (i.e., a counter-evidence to the inescapability condition). In the case of wine, you just need basic knowledge of what wine is and/or why you value the wine you like. In more general terms, when you try to approximate an ideal, the aspect you focus on must be relevant in the sense that it captures a part of quality that contributes to the idealness of the ideal.

Per the prevalence condition, it also weakens the anti-approximation argument that many examples are detached from real-world politics. A more relevant example is discussed by Wolff. He warns that a partial implementation of Dworkin's theory of equality of resources may lead to Thatcherism. This is because real-world policymakers will likely focus on the element of 'responsibility for choice', which is easier to implement than broader background conditions (Wolff 2019, p. 16). However, similarly to the wine case, the problem can be avoided if one has a proper understanding of Dworkin's normative theory. A foundational value in Dworkin's theory is equal concern for every person, which is violated by laissez-faire economic policies (Dworkin 2011, p. 2). Individual responsibility for choice is valued in so far as it is an expression of equal concern for the dignity of every person. Forcing people to accept responsibility for something which they have no moral responsibility for is the opposite of what Dworkin's theory intends. Knowing this, trying to approximate the Dworkinian ideal society by making individuals take responsibility for arbitrary market outcomes would be a silly mistake indeed. Wolff's criticism highlights the importance of articulating what is important in the normative ideal and why, but it does not present an insurmountable problem to approximation.

The second explanation of how the problem of second best may arise concerns inter-dimensional interactions. Different aspects of the ideal may be closely interconnected such that a change in one affects other(s). Goodin cites the interactions between education, health and employment as an example. Policymakers need to pay attention to all these aspects holistically rather than separately (Goodin 2012, p. 159). Estlund in effect qualifies Goodin when he says what is problematic is 'strong negative synergy' (Estlund 2020, p. 286). It refers to phenomena where the effect of an inter-dimensional interaction is hugely undesirable, and, thus, adding more elements that are part of the ideal may worsen a non-ideal situation rather than improve it. This happens because some element of an ideal has positive value only in some specific combination with other elements. Placed in a different context, the same element may have a negative value. Estlund employs the example of pills (Estlund 2020, p. 274). When you are prescribed three pills, taking two out of the three may make the situation worse than taking no pills. The lesson is that some contributing elements in an ideal need to be treated as an inseparable unit or a 'system' (Estlund 2020, p. 286).

Escaping the problem of inter-dimensional interaction may be practically difficult, but it is not unavoidable. Fundamentally, it is the same problem as the selection of the aspect or dimension for approximation. The solution is also the same: you just need to have an accurate understanding of the right unit of analysis. Henceforth, I will use the term *frame of analysis* to refer to the aspect and unit of analysis for approximation.

Focusing for now on unit of analysis, let us discuss Goodin's and Estlund's examples. In those cases, since you know how the problematic interactions/synergies may happen, avoiding them should be possible. You just need to bundle those interacting elements, for example, by always considering education, health and employment together.[12] Put more generally, the solution is to arrange the unit of analysis so that there is no undesirable interaction between the dimension or aspect along which you attempt to approximate the ideal. This may seem like an artificial manipulation, and it is, but the units of analysis originally used, such as education and pills, are already human constructs.[13]

Thus, it is far from conclusive that this way of explaining the problem of second best commands a substantive enough warning against approximation. While the inescapability of inter-dimensional interaction seems somewhat plausible, its prevalence is relatively unclear as the theorists cite few examples of public policy, especially in the form of strong negative synergy. More arguments or to-the-point examples are needed to support the moderate warning.

Let us consider the third explanation of how the problem of second best might afflict the strategy of approximation: incomplete ranking. It is a difficulty whereby the ideal fails to rank all of the relevant options. If an evaluation of options is partially mistaken, an attempt to approximate an ideal would also be unreliable. Wiens (2015) argues that ideal principles have too limited a range of application for ranking options in the real world. However, Wiens does not explain why this is so. Instead, he appeals to an example of Rawls's ideal theory. Wiens compares two worlds, w_1 and w_2. Both worlds satisfy equal basic liberties and fair equality of opportunity, but in w_1 the difference principle is applied, and the position of the least advantaged is slightly better at the expense of greater inequality than in w_2 where the difference principle is not applied (Wiens 2015, p. 443). When you use Rawls's ideal theory as a benchmark, it ranks w_1 higher than w_2 despite much greater inequality in w_1 because w_1 is closer to the ideal. Wiens says this is 'deeply counterintuitive' in light of the core values in Rawlsian theory itself, demonstrating the inability of Rawlsian ideal theory to rank all relevant options (Wiens 2015, p. 443).

This is not a compelling example. Ranking w_1 higher than w_2 is compatible with Rawls's theory (or one natural reading of it) for two reasons. First, the position of the least advantaged is measured in terms of expectations in the index of primary

[12] Analyses of policy or politics naturally requires combining multiple dimensions (Bache et al. 2016).

[13] Changing units of analysis can be practically difficult. For example, creating an institutional framework for overseeing education, health and employment may be politically infeasible in the short term. Also, if the interactions and synergies between different elements cannot be predicted, there is naturally no way to avoid them pre-emptively.

goods, including social bases of self-respect, which is said to be the most important component (Rawls 1999, p. 348, pp. 478–479). This means that the least advantaged in w_1 is expected to have a greater amount of primary goods, with the social bases of self-respect being the weighty component, than the least advantaged in w_2. In this sense, the position of the least advantaged in w_1 is meaningfully improved compared to the one in w_2, even if the difference may seem small. Second, Rawls's theory is not concerned with the size of inequality itself (Rawls 2005, p. 283). What matters instead is fairness of social institutions in the sense that they are arranged with the maximum benefit of the least advantaged in mind. Inequality is a problem of justice in so far as it hinders the realisation of principles of justice that are deemed fair to all citizens. Thus, it makes sense that Rawls's theory ranks w_1 above w_2 Wiens's broader purpose seems to be to defend the position I call case-by-case empiricism. However attractive that position may be (see the penultimate section), we cannot say that the problem of incomplete ranking provides credible support for the moderate warning.

To summarise the discussion so far, evidence to support the moderate warning seems insufficient. It cannot be maintained even as reliable speculation. To that extent, the anti-approximation case is inconclusive.[14] It leaves room for defending the approximation of an ideal as a sensible default strategy of action guidance in the real world.

Sophisticated Approximation

This section develops a positive case for the approximation of an ideal as a strategy of action guidance. Building on the foregoing discussion, I develop the account of sophisticated approximation. While admittedly broad-brush, it goes some way towards showing conditions for successful approximation, with reference to beneficial variables: desirability, feasibility of the ideal and knowledge of the ideal.

Proposal

How can we envision a successful move towards the realisation of the ideal through approximation? I propose carefully selecting a frame of analysis and reforming social institutions by degrees, where and when it is feasible, so that they become increasingly similar to the ideal over time. Simultaneously, as we proceed, updated knowledge is used to improve the process itself. Call it *sophisticated approximation*.

This approach has three basic features. First, careful selection of the good frame of analysis (i.e., aspect and unit of analysis) is crucial. Taking an extreme case, if you can identify a single dominant aspect, approximation will be quite straightforward. For example, when you are trying to raise a million dollars, the state that is descriptively closer to the target is highly likely to be more desirable. Of course,

[14] The weak warning still has some anti-approximation implications, a part of which is discussed in the penultimate section.

there are usually several important aspects, and you need to pay attention, to the best of your ability, to all of the aspects and interactions between them. Also, the unit of analysis may need to be adjusted. For example, instead of treating health as a dimension that is independent from education and employment, you may need to expand the dimension of health to include factors related to education and employment based on your understanding of the social determinants of health. The aim of such an adjustment is to account for significant and systematic interactions between what were originally conceived as independent dimensions. Second, the validity of approximation is judged by its incremental performance rather than by a single-shot outcome. All of the examples of the problem of second best considered so far focus on a single-shot approximation. However, in the context of social institutions, approximation of an ideal normally takes several incremental steps. A long-term tendency to improve the institutions is typically more important than the immediate outcome of a particular attempt at reform because ideal social institutions cannot be realised in the short term. Third, this approach is not committed to always adopting approximation as a tactic. When you expect approximation to perform significantly badly, you should adopt another tactic for that moment. What matters is that you pursue the approximation of an ideal with a reasonable degree of consistency throughout multiple incremental attempts over time.

By way of elaborating the first aspect further, I propose two stages of sophistication. The initial stage is an easy target whose effect may be limited to avoiding particularly bad cases of approximation. But this already makes approximation much better than portrayed by the critics. The second stage involves more deliberate examinations and aims to make approximation perform even better.

To explain the first stage, I need to introduce the notions of attributional and relational commonalities[15] as different types of similarity that feature in our similarity judgements. While non-relational, attributional commonality considers whether a certain attribute (e.g., colour) is shared between the objects being compared, relational commonality looks at relations or patterns of such attribute(s) within each of the objects being compared and considers whether there is any similarity in patterns or relations between them. Relational commonality can be understood by thinking of two objects, each consisting of three parts. All parts of object 1 are yellow, while all parts of object 2 are blue. There is relational commonality between the two objects because both exhibit internal colour consistency. Another example of relational commonality is the similarity between the solar system and an atom (Medin et al. 1993, p. 257). Despite having no attributional commonality, the two are similar in their structural features, such as having a large core with smaller parts orbiting around it. A judgement of relational commonality *can* be complex when several attributes are involved, but it need not be. It is a part of our ordinary judgement. Most people understand the similarity between the solar system and an atom without difficulty.[16]

[15] Similar distinction is also expressed by the terminology of horizontal versus vertical relations (Bartha 2010, 2019; Schonen 2022).

[16] A judgement of similarity focusing on the whole rather than a specific part of objects is not necessarily more advanced (Smith 1989).

Now, the first stage of sophistication for our selection of frames of analysis is simply to make some consideration of relational commonality in addition to attributional commonality. At this stage, sophistication does not require anything beyond applying ordinary judgement to thinking about the similarity in terms of relational commonality. Nevertheless, merely adding such consideration is a significant step forward from the crude approximations depicted in the critics' examples we saw above, where no attention is paid to relational commonality. In the wine case, for example, we should consider a balance of tannins, acidity, sweetness, finish length, etc., and not just whether some such attributes are present. Even assuming that our judgement of relational commonality is not particularly good, it is reasonable to assume that we will end up with a better frame of analysis than when we make no such consideration.

The second stage of sophistication involves a more calculated assessment of frames of analysis. Here, I understand a good frame of analysis as the one where descriptive proximity correlates with desirability (I will shortly add two more correlations).[17] Let us call this *proximity-desirability correlation*. The amount of money raised is a frame of analysis with a high proximity-desirability correlation for the fund-raising project. If we have reason to believe that the correlation is sufficiently high, then the approximation of the ideal may be regarded as a useful guide. Let me illustrate: Let us assume that American citizens agree that the German co-determination model is their ideal economic system, although there is no prospect of fully replicating it. Assume further that the proximity-desirability correlation is estimated to be roughly 0.7. The citizens undertake a series of reforms that incrementally make US institutions look more like the German model. An individual piece of reform may fail to realise a more desirable institutional arrangement than before, as the correlation is not perfect; however, the citizens can reasonably expect that the series of reforms over time will result in a significant improvement in the desirability of their institutions. In contrast, if the proximity-desirability correlation is low or negative, it is destructive to the incremental approximation of the ideal institution.

Of course, it is extremely difficult to know the rate of a proximity-desirability correlation. However, even our common-sense judgement can distinguish particularly bad frames and frames that are reasonably likely to be relevant.[18] For the German-style economic system, the bargaining power of workers seems important, while the culture of punctuality is probably not, and speaking the German language is definitely not. Beyond common sense, one important epistemic guideline is to look for a frame of analysis that is based on knowledge of structural relations regarding the objects being compared. Investigating a practical epistemic warrant for identifying reliable and informative similarity arguments, Tom Schoonen (2022) proposes

[17] Desirability of a state and desirability of a path to the state can be distinguished conceptually. Both of them matter, and they may be closely connected in cases of small incremental changes.

[18] Selecting a relevant frame of analysis based on a given context is an ordinary feature of our similarity judgement (Medin 1993; Goldstone and Son 2005). Also, a major role played by our ordinary knowledge in policymaking processes should not be underestimated (Lindblom and Cohen 1979).

the requirement of 'explicit structural knowledge'.[19] Explicit structural knowledge tells us about loosely causal relations between certain descriptive features and some desirable variable (e.g., monetary value, or normative desirability) within the target object of comparison (Schoonen 2022, p. 8). It refers to an abstract system of knowledge that provides some explanation of loosely causal nature beyond mere correlations without being specifically about concrete cases (Schoonen 2022, p. 12).

This requirement of explicit structural knowledge is in line with our earlier discussion about how theoretical knowledge about the ideal and empirical knowledge about inter-dimensional interactions should help select a relevant frame of analysis. When the justice of social institutions is in question, normative theories can be useful not only in understanding what the ideal is like but also in selecting relevant frames for comparing different institutional arrangements (Gilabert 2017, p. 118). Knowledge of normative theories helps us spot a crude form of approximation and choose a more reliable frame of analysis that has a reasonably high proximity-desirability correlation. Some examples of explicit structural knowledge in Rawls's theory are the lexical priority of the principles, constitutional constraints on forms and power of democratic legislature and institutions of property-owning democracy. So, institutional features that are based on or consistent with such theoretical knowledge are good candidate features to focus on in sophisticated approximation. Similarly, theoretical knowledge about Dworkin's ideal should help us effectively approximate the Dworkinian ideal without unwittingly creating Thatcherite institutions. Examples of such knowledge include the importance of equal respect for persons and the necessary alignment between the level of individual responsibility and the individual's genuine ability of choice.

But how do we actually use such knowledge to evaluate frames of analysis? There is no simple formula, but one heuristic method may be to ask why you value the feature you intend to focus on and see if your knowledge offers a reasoned justification. Repeat the question by changing the aspect and unit of analysis and give weights to all the reasons provided. This gives you a set of possible frames of analysis with different weights. You should eliminate low-ranking frames of analysis and use the weightier frame(s).[20] For example, you can use general knowledge about society to assign a much greater weight to labour bargaining power than to punctuality when evaluating economic systems. Also, you might use theoretical knowledge about Rawls's lexical priority to give greater weight to policies justified by the consideration of the basic minimum than those justified (solely) by the difference principle.[21]

Finally, the more reform we implement pursuant to the ideal, the more desirable the next piece of the reform is likely to be. This is because of institutional complementarities, by which similar institutions tend to reinforce each other (Sirsch 2020,

[19] Schoonen's argument is based on the model of similarity argument developed by Bartha (2010, 2019).

[20] For a method of similarity assessment involving aspects with different weights, see Smith and Osherson (1989).

[21] Somewhat similar reasoning can be found in the judicial practice of *stare decisis* where both theoretical and specific empirical knowledge are used to decide whether the present case is relevantly similar to the precedent. *Stare decisis* is also similar to sophisticated approximation in that it is justified by its long-term performance despite occasional costs in justice (see Bartha 2010, pp. 246–248).

pp. 98–101). So, if the proximity-desirability correlation is low, it tends to improve as we incrementally approximate the ideal.

Correlation with Feasibility and Knowledge of the Ideal

I would now like to consider two additional variables that may correlate with the proximity of institutional design: feasibility and knowledge of an ideal. Considering three correlations will help us identify effectively good frames of analysis where approximation can be considered worthwhile.

By feasibility of an ideal, I mean the likelihood that a transformation of the status quo realises the ideal.[22] When we incrementally approximate our social institutions to the ideal, such change may make the full realisation of the ideal more feasible in this sense. Call this the *proximity-feasibility correlation*.

This correlation is distinct and independent from the proximity-desirability correlation. For one thing, the proximity-feasibility correlation is more plausible than the proximity-desirability correlation. When B is closer to A than C is, it is usually easier to get to A from B than from C. It is also noteworthy that the proximity-feasibility correlation may hold in cases where the proximity-desirability correlation does not. Suppose you rank desirability of social welfare schemes in the order of: (1) the egalitarian Universal Basic Income (UBI); (2) the German-style social state; and (3) the American-style Negative Income Tax (NIT), where the UBI is structurally more similar to the NIT than to the German social state. Here, there *could* be a higher feasibility of transforming the NIT into the most desirable UBI scheme than transforming the second most desirable German social state into the UBI. If this is the case, the descriptive proximity of each option to the best option correlates with the feasibility of the best option but not with the desirability of each option (the proximity-feasibility correlation is positive, but the proximity-desirability correlation is negative). Thus, even where the proximity-desirability correlation is low or negative and there is a good chance that a reform to approximate the ideal would make social institutions morally inferior to the status quo (e.g., a move from (2) to (3)), you might be reasonably confident that the same reform would improve the feasibility of achieving the ideal because of a high proximity-feasibility correlation.

Of course, proximity-feasibility correlation is not always high. To use the earlier example about wine, turning the mix of the red and the white into the pure red is difficult. Comparatively, turning red grapes (with skin) into red wine may prove easier, although grapes look a lot less like red wine than the pink liquid does. Hence, there is no correlation between descriptive proximity and feasibility. Again, though, the frame of analysis is important. Under some frame such as material composition, red grapes may be more similar to red wine than the red-and-white blend is. Therefore,

[22] Note that this is different and independent from the feasibility of a measure that approximates the ideal.

by selecting a good frame of analysis, we can say that the proximity-feasibility correlation may hold even in this case.[23]

Additionally, there is a basis for believing that proximity-feasibility correlation tends to hold for institutions, namely, path-dependency. Path-dependency means that the existing institution tends to make similar institutions easier to realise and dissimilar ones harder to realise (Sirsch 2020, pp. 98–101). Thus, when extant institutions share few features with the ideal, a reform in the direction of the ideal can be difficult to implement. But, crucially, the more the extant institutions become similar to the ideal, the easier it becomes to implement additional reforms in the direction of the ideal.

The other variable that may correlate with proximity is knowledge about the ideal. We tend to gain a greater understanding of an ideal social institution as we move closer to it.[24] When you have something that is more similar to the ideal, you can learn more about the ideal because of the greater commonality of features. Since we usually do not have a detailed understanding of ideal social institutions, any opportunity to learn something about them through their approximation can be an essential source of information. Hence, we can think about the *proximity-knowledge correlation*.

Like the proximity-feasibility correlation, the proximity-knowledge correlation is never perfect, and the rate of correlation changes depending on what aspect you focus on. For example, if you approximate the German economic system by focusing on punctuality, you will probably learn nothing of substance about it. And, knowledge of the ideal is distinct from the immediate desirability of the option and the feasibility of the ideal, though knowledge may be used to increase the feasibility.

The different stages of sophistication discussed above also apply to feasibility and knowledge. One difference is that empirical knowledge may be more important than theoretical knowledge for achieving the second stage of sophistication for the proximity-feasibility correlation. For the proximity-knowledge correlation, since both theoretical and empirical knowledge about the ideal are needed, available knowledge of the both kinds are important for identifying a frame of analysis with a reasonably high correlation. Additionally, regarding knowledge of the ideal, we can expect a positive feedback loop where knowledge gained through the process of approximation can be used for further improving the rate of three correlations for the following iterations.[25] This means that we can often start with satisfying the initial stage of sophistication and move to satisfying the second stage of sophistication, gradually improving the rates of the three correlations on the way.

[23] It is worth noting that even with a bad frame of analysis, a proximity-feasibility (or desirability, for that matter) correlation might still be positive and roughly linear: keep adding red wine to a red-and-white mix (put in a huge container), and it will become increasingly similar to pure red wine until eventually there is no perceptible difference.

[24] This seems rather robust. Gaus (2016) expresses the same thought in a negative way by the term 'neighborhood constraint'.

[25] Some researchers of psychology of similarity argue that surface similarities that are ultimately abandoned can still be useful as heuristic for understanding deeper similarity relations (Medin and Ortony 1989). Something similar may be the case with improving our knowledge of ideal social institutions.

There are two further implications of considering feasibility and knowledge on top of desirability. First, if an attempt to approximate an ideal has a sufficiently high correlation with two or three variables rather than one, the move can be considered more valuable. In that case, we have a better reason to attempt to approximate the ideal. Second, an approximation may still be considered valuable even if it fails to bring about a more desirable state than the status quo, if the proximity-feasibility correlation and/or proximity-knowledge correlation is sufficiently high. As long as the proximity-feasibility correlation and/or proximity-knowledge correlation has a reasonably high rate, we can at least move towards a state where the ideal is more feasible and/or we can learn more about the ideal. To use the example of Dworkin's theory, it *could* be the case that the bad approximation that leads to Thatcherism nonetheless makes it easier to realise Dworkinian ideal institutions and/or improve our knowledge about that ideal. If those gains in feasibility and knowledge are great enough to offset the loss in desirability from the immediate change, it could be considered a sensible attempt to approximate the Dworkinian ideal.[26] My account of sophisticated approximation helps us conceptualise such trade-offs if not resolve them.

While details remain to be filled in, I believe that these considerations provide good reasons for claiming that the strategy of sophisticated approximation is both practical and presumptively reliable over time. My strategy is to pursue greater descriptive proximity to the ideal based on the selection of good frames of analysis with eyes on three beneficial correlations. It can start with simply utilising our ordinary judgement and work towards a further sophistication of the process of approximation itself. We should remain cautious, but such a strategy can be considered reasonably reliable until we have reason to suspect otherwise. And, if you foresee a considerably high risk of a major failure in a particular iteration of the incremental approximation, you should opt for a better path—and this does not mean the general strategy of sophisticated approximation has failed. My claim is that approximation is sufficiently sensible as a default strategy, not that you should stick to approximation no matter what. But this is still a strategy of approximation because the course of social reform is guided by the general expected reliability of descriptively approximating the ideal rather than by the desirability of a specific piece of reform (or state of affairs) at every instance. The next section considers an objection and alternatives by way of further articulating the merit of sophisticated approximation.

Objection and Alternatives

We should consider the possibility that there is a better strategy of action guidance that replaces approximation not just temporarily but as a default strategy. An alternative should take over as a default strategy if it is expected to outperform sophisticated approximation. After addressing a concern about the demandingness of

[26] Perhaps desirability should be given greater weight than feasibility or knowledge of the ideal. Still, situations can arise where gains in feasibility and knowledge can make up for a loss in desirability.

sophisticated approximation, I consider three candidate strategies: the robustness strategy, countervailing deviation and case-by-case empiricism. These are endorsed by the prominent critics of approximation mentioned in this paper and could be viewed as an alternative to sophisticated approximation. But I argue that they are subsumed under the strategy of sophisticated approximation, more local or otherwise less attractive than it.

Demandingness Objection

The reader might worry that my attempt to save approximation made it too complicated. In particular, the second stage of sophistication for selecting good frames of analysis may seem demanding. Wolff's example about pseudo-Dworkinian policy is important here. For how can we expect policymakers to read through *Sovereign Virtue*? A similar epistemic demandingness exists for empirical matters, too. Policymakers make errors in empirical assessments more often than we hope, even with the assistance of scientific advisors. Maybe we just cannot expect policymakers to have the epistemic capability for selecting good frames of analysis.

This is a real concern. But I wish to quell the worry by pointing out that sophisticated approximation is not particularly demanding, considering what serious policymaking must involve anyway. Perhaps policymakers deserve more credit. But more importantly, the difficulty is not unique to the approximation of an ideal. Any policymaking requires careful empirical assessment, normative judgement, fine-tuning and recalibration. Also, we need to evaluate sophisticated approximation against alternatives (see the following subsections).

Meanwhile, there are reasons to be modestly optimistic about obtaining good frames of analysis. Recall that a good frame of analysis is one where the three virtuous correlations are sufficiently high. First of all, the rates of correlations do not have to be very high. Ultimately, they just have to beat coin-tossing. For, as long as the rates of correlations are positive, your pursuit to approximate an ideal moves in the right direction over time, albeit slowly and with occasional setbacks. Also, a low correlation in one variable may be compensated by a sufficiently high correlation in other variable(s). We saw that the proximity-feasibility and proximity-knowledge correlations are both likely to be higher than the proximity-desirability correlation; and even the proximity-desirability correlation tends to become higher as the incremental pursuit of the ideal moves forward. I also discussed how we can start with the easier task of satisfying the initial stage of sophistication, with the expectation of positive epistemic feedback supporting all three correlations.

Robustness Strategy

Goodin proposes what he calls the robustness strategy in light of the institutional path-dependency and possibility of unexpected change in feasibility constraints, where a policy currently considered unavailable may become available in future. For example, think of the choice between the third-best and fourth-best options, which are the only available options at certain timepoint (call it T_1). Approximation would

recommend the third-best; but, it may be that the fourth-best is more amenable to adopting the second-best option should it become available later (T_2) (Goodin 2012, pp. 170–171). Contra approximation, this fact counts for choosing the fourth-best option at T_1.

Based on these considerations, the robustness strategy prescribes you to '[c]hoose a policy that is robust against changes in feasibility' (Goodin 2012, p. 171). That is, when you evaluate a policy, you need to consider how its implementation may constrain available courses of action in the future, in addition to its immediate effect. Thinking ahead in this way is important because you cannot make a new policy every time the feasibility constraints change.

While Goodin's observation is sound, the robustness strategy cannot be a better alternative to sophisticated approximation for two reasons. First, the robustness strategy is itself a form of approximation. Although its time frame is longer than a simple approximation, the robustness strategy does not defy approximation. Notice that, Goodin's puzzle exists only if the approximationist is confined to thinking only about the desirability in T_1. But nothing bars the strategy of approximation from adopting the time frame covering T_1 and T_2, and the robustness strategy is doing just that. The choice of the fourth-best option at T_1 is preferred precisely because you want to approximate the ideal at T_2. So, robustness strategy simply pursues an approximation of the ideal in a longer term. Second, the consideration of future options is already included in the attention to the proximity-feasibility correlation. A frame of analysis should be chosen with the proximity-feasibility correlation in mind so that incremental steps towards the realisation of the ideal are as secure as possible. As considerations of future options are accommodated in sophisticated approximation, it subsumes the robustness strategy.

Case-by-Case Empiricism

Another alternative is what might be called *case-by-case empiricism*. It capitalises on the weak warning that the approximation of an ideal does not *necessarily* realise the optimal situation. It might, but you can never tell without empirical knowledge about the specific situation. You should simply evaluate options through empirical investigations every time you face a new situation and select the most desirable one. Sen, Goodin and Wiens make arguments to this effect (Sen 2009; Goodin 2012; Wiens 2015). This position poses a threat to sophisticated approximation because it tells you to adopt *no* default strategy; you must evaluate every option afresh without presumption.

Could case-by-case empiricism be a better position than sophisticated approximation? I believe not. Defiance of presumption may not be a virtue for real-world policy guidance, and to err on the side of caution may prove rather costly. I believe sophisticated approximation performs better on three counts.

First, sophisticated approximation provides guidance with greater consistency and predictability than case-by-case empiricism. Case-by-case empiricism may make one recommendation ('introduce UBI') at one time point and the complete opposite ('abolish UBI') at the next, if each policy is judged most desirable in each

situation. No attention is paid to the consistency or continuity of policies. To be sure, consistency may not itself be worth going after. However, by aiming to move towards the ideal, recommendations of sophisticated approximation across time exhibit greater consistency and predictability as a useful by-product.[27]

Second, sophisticated approximation is less prone than case-by-case empiricism to be pulled away from the ideal or trapped in a local optimum far inferior to the ideal. Case-by-case empiricism is more likely to lead away from the ideal because it is solely (or primarily) guided by the desirability of the immediate policy change. Sophisticated approximation may encounter similar difficulties, but the risk is smaller because it pursues the approximation of the ideal—sometimes even at the expense of desirability, provided that the gains in knowledge and/or feasibility of the ideal are great enough.

Third, evaluating every option with no theoretically informed orientation or prioritisation of options may be very costly. The less you assume in advance, the harder it is to assess options. The case-by-case empiricist cannot decide until *every* option is considered, for no end state is prioritised. Meanwhile, you let the existing institutions run their own course. Comparatively, sophisticated approximation economises the process *because* it presumes a long-term direction and the first option to be considered. In other words, the project of approximation can be a useful presumption if reasonable care is taken to formulate it (as I outlined in this paper). So, whatever benefit cautious empiricism brings, its cost need also be recognised.

Countervailing Deviation

Finally, let us consider Estlund's countervailing deviation. After pointing out the fallibility of approximation of an ideal, Estlund introduces countervailing deviation as an alternative way to use an ideal:

> Sometimes when there is a missing element from some valuable or even ideal condition A, we can ask what value is thereby lost in that condition B: what kind, and how much. […] [S]ometimes an appropriate countervailing measure might be to deviate even further from the structural model or ideal. That would make the resulting situation, call it C, less similar to the ideal scenario A than B is. Even so, in principle, the second subtraction—the further departure—could, in principle, partly or fully restore the lost value. (Estlund 2020, pp. 290–291)

While moving away from the ideal, countervailing deviation uses the ideal scenario as a model through which to understand the conditions required for the value in question to exist. Estlund explains the example of campaign finance restriction as a deviation from the ideal of free speech (Estlund 2020, pp. 296–300). It is a deviation because it involves a levelling-down of the amount of speech rather than an increase. Such a restriction can be justified as a way to restore the fair value

[27] On the importance of consistency, see North (2017).

of political liberty, which is lost by unequal access to the wherewithal of speech. According to Estlund, countervailing deviation can effectively utilise the ideal to guide our action when approximation of the ideal misleads us (Estlund 2020, pp. 302–303).

However, I do not believe countervailing deviation is fit to be an alternative general strategy of action guidance in the real world. For one thing, I suspect countervailing deviation has limited real-world applications. Presupposing commitment to an ideal, it can certainly happen that you better deviate from (approximation of) it. However, what motivates such a move seems normally to be a revision or an update of the ideal. It seems rare to find cases of countervailing deviation that are not *more straightforwardly* an approximation or realisation of a somewhat different ideal. Campaign finance restriction is no exception. It can more straightforwardly be understood as a pursuit of the ideal (or model) of the fair value of political liberty or balanced active speech. Estlund himself seems to admit this (Estlund 2020, p. 291), and it may not matter theoretically. But as a guide to real-world policymaking, countervailing deviation seems rather elusive.

Furthermore, countervailing deviation does not have to deny the general validity of sophisticated approximation. Insofar as countervailing deviation is not inconsistent with giving weight to the descriptive proximity to the ideal,[28] it can be subsumed under or embedded in the strategy of sophisticated approximation. In this case, countervailing deviation is a local tactic to be used temporarily when a straightforward approximation looks unpromising, which does not abandon the project of sophisticated approximation.

Conclusion

This paper has defended the approximation of an ideal as an approach of action guidance on public policy and social institutions in the context of the real world. Granted that approximation faces difficulties, as illustrated by the problem of second best, the difficulties can be mitigated by sophistication of the project. We identified some conditions where approximation can be a sufficiently reliable guide over time. Also, sophisticated approximation is practical enough a project, though more work is needed on details.

My argument against the alternatives is not meant to show that approximation is the unrivalled champion. In particular, we can foresee the case-by-case empiricists

[28] This way of understanding countervailing deviation makes intuitive sense. If there are two ways to countervailingly deviate that restore the same value to the same degree, but one realises greater descriptive proximity to the ideal than the other, this strategy recommends the first option. However, the idea of countervailing deviation detached from approximation cannot explain this choice. Alternatively, countervailing deviation may be understood to deny any epistemic or heuristic value to descriptive proximity. In this case, countervailing deviation would use only the full ideal as a source of knowledge just for ranking options in terms of immediate desirability. That would essentially be a case-by-case empiricism with the added epistemic requirement of investigating the abstract realm of pure values. My objections to case-by-case empiricism would apply to this latter understanding of countervailing deviation.

responding to my challenge with sophistication of their approach—though such a move would likely make case-by-case empiricism even more complex and costly. What I hope to have shown is that approximation is a promising strategy of policy guidance, and until we have a clear outperformer it remains one of the best.

I should add that key insights of case-by-case empiricism are appreciated. Assessing desirability of a policy based on empirical knowledge of a changing situation is an important part of sophisticated approximation. Moreover, after considering empirical evidence, the sophisticated approximationist may decide to pursue desirability of an immediate policy change at the expense of proximity to the ideal. Such empiricist-like thinking may even be dominant across a range of situations that are plagued by grave injustice or imminent crisis (e.g., institutional slavery). Here, making some immediate improvement in the desirability of the existing social institutions may be overwhelmingly important.

Even in such situations, however, it matters that the approximation of the ideal is held as a default strategy. Approximation of the ideal should be the first option to be considered, and, while the pursuit of approximation can be temporarily halted, you should return to it once it becomes sensible; that is, once you can reasonably expect the aggregate of the three virtuous correlations for the foreseeable future to be sufficiently high.

Acknowledgements I am grateful for their helpful comments to Shai Agmon, Rufaida Al Hashmi, Geertje Bol, Jaime Draper, Gen Fukushima, Hayatora Hotta, Shu Ishida, Takuto Kobayashi, Elsa Kugelberg, Masaya Miyamoto, Ryo Ogawa, Jun-ichi Saito, Zosia Stemplowska, Stuart White, Jo Wolff and two anonymous reviewers.

Declarations

Conflict of interest The author declares that he has no conflict of interest.

References

Arvan, Marcus. 2019. Nonideal justice as nonideal fairness. *Journal of the American Philosophical Association* 5 (2): 208–228.

Bache, Ian, Ian Bartle, and Matthew Flinders. 2016. Multi-level governance. In *Handbook on theories of governance*, ed. Christopher Ansell and Jacob Torfing, 486–498. Cheltenham: Edward Elgar.

Bartha, Paul. 2010. *By parallel reasoning*. Oxford University Press.

Bartha, Paul. 2019. Analogy and analogical reasoning. In *The Stanford encyclopaedia of philosophy*, ed. E. N. Zalta. https://plato.stanford.edu/entries/reasoning-analogy/#PriJus. Accessed 18 April 2022.

Dworkin, Ronald. 2011. *Justice for hedgehogs*. Cambridge, MA: Harvard University Press.

Estlund, David. 2011. Human nature and the limits (if any) of political philosophy. *Philosophy and Public Affairs* 39 (3): 207–237.
Estlund, David. 2020. *Utopophobia: On the limits (if any) of political philosophy*. Princeton, NJ: Princeton University Press.
Gaus, Gerald. 2016. *The tyranny of the ideal: Justice in a diverse society*. Princeton, NJ: Princeton University Press.
Geuss, Raymond. 2016. *Reality and its dreams*. Cambridge, MA: Harvard University Press.
Gilabert, Pablo. 2017. Justice and feasibility: A dynamic approach. In *Political utopias: Contemporary debates*, ed. Michael Weber and Kevin Vallier, 95–125. Oxford: Oxford University Press.
Goldstone, Robert, and Ji Yun Son. 2005. Similarity. In *Cambridge handbook of thinking and reasoning*, ed. Keith Holyoak and Robert Morrison. Cambridge: Cambridge University Press.
Goodin, Robert. 1995. Political ideals and political practice. *British Journal of Political Science* 25: 37–56.
Goodin, Robert. 2012. The bioethics of second-best. In *Global justice and bioethics*, ed. Joseph Millum and Ezekiel J. Emanuel, 153–177. Oxford: Oxford University Press.
Guillery, Daniel. 2021. The concept of feasibility: A multivocal account. *Res Publica* 27: 491–507.
Hamlin, Alan. 2017. Feasibility four ways. *Social Philosophy and Policy* 34 (1): 209–231.
Hamlin, Alan, and Zofia Stemplowska. 2012. Theory, ideal theory and the theory of ideals. *Political Studies Review* 10: 48–62.
Lindblom, Charles, and David Cohen. 1979. *Usable knowledge: Social science and social problem solving*. New Haven, CT: Yale University Press.
Lipsey, Richard, and Kelvin Lancaster. 1956. The general theory of second best. *The Review of Economic Studies* 24 (1): 11–32.
Medin, Douglas, and Andrew Ortony. 1989. Psychological essentialism. In *Similarity and analogical reasoning*, ed. Stella Vosniadou and Andrew Ortony, 179–195. Cambridge: Cambridge University Press.
Medin, Douglas, Robert Goldstone, and Dedre Gentner. 1993. Respects for similarity. *Psychological Review* 100: 254–278.
Mills, Charles. 2017. *Black rights/white wrongs: The critique of racial liberalism*. New York: Oxford University Press.
North, Richard. 2017. Principles as guides: The action-guiding role of justice in politics. *The Journal of Politics* 79 (1): 75–88.
Rawls, John. 1999. *A theory of justice: Revised edition*. Cambridge, MA: Harvard University Press.
Rawls, John. 2005. *Political liberalism: Expanded edition*. New York: Columbia University Press.
Schoonen, Tom. 2022. Possibility, relevant similarity, and structural knowledge. *Synthese* 200 (39): 1–22.
Sen, Amartya. 2009. *The idea of justice*. Cambridge, MA: Harvard University Press.
Simmons, A. John. 2010. Ideal and nonideal theory. *Philosophy and Public Affairs* 38 (1): 5–36.
Sirsch, Jürgen. 2020. *Designing realistic utopia: Ideal theory in practical political philosophy*. Baden-Baden: Nomos.
Smith, Linda. 1989. From global similarities to kinds of similarities: The construction of dimensions in development. In *Similarity and analogical reasoning*, ed. Stella Vosniadou and Andrew Ortony, 146–178. Cambridge: Cambridge University Press.
Smith, Edward, and Daniel Osherson. 1989. Similarity and decision making. In *Similarity and analogical reasoning*, ed. Stella Vosniadou and Andrew Ortony, 60–75. Cambridge: Cambridge University Press.
Stemplowska, Zofia. 2016. Feasibility: Individual and collective. *Social Philosophy and Policy* 33 (1–2): 273–291.
Swift, Adam. 2008. The value of philosophy in nonideal circumstances. *Social Theory and Practice* 34 (3): 363–387.
Volacu, Alexandru. 2018. Bridging ideal and non-ideal theory. *Political Studies* 66 (4): 887–902.
Wiens, David. 2015. Against ideal guidance. *The Journal of Politics* 77 (2): 433–446.
Wiens, David. 2016a. Motivational limitations on the demands of justice. *European Journal of Political Theory* 15 (3): 333–352.
Wiens, David. 2016b. Assessing ideal theories: Lessons from the theory of second best. *Politics, Philosophy & Economics* 15 (2): 132–149.
Wolff, Jonathan. 2015. *An introduction to political philosophy*, 3rd edn. Oxford: Oxford University Press.

Wolff, Jonathan. 2019. Method in philosophy and public policy: Applied philosophy versus engaged philosophy. In *Routledge handbook of ethics and public policy*, ed. Annabelle Lever and Andrei Poama, 13–24. London: Routledge.