



# Too many cooks

Alex Horne<sup>1</sup> 

Received: 31 March 2021 / Accepted: 29 June 2022 / Published online: 21 July 2022

© The Author(s) 2022

## Abstract

The existing literature on the rational underdetermination problem often construes it as one resulting from the ubiquity of objective values. It is therefore sometimes argued that subjectivists need not be troubled by the underdetermination problem. But on closer examination, it turns out, they should. Or so I will argue. The task of the first half of this paper is explaining why. The task of the second half is finding a subjectivist solution the rational underdetermination problem. The basic problem, I argue, is as follows. Idealizing subjectivism generates too many ideal selves to deliver determinate or commensurable options regarding what non-ideal deliberating agents ought to do. My solution: these idealized options should be assessed from the only perspective we can, in fact, occupy, namely, that of our non-ideal, actual selves. Deciding what to do therefore becomes, in part, an exercise in deciding who to be. But one might now worry this just moves the arbitrariness bump in the rug. Privileging the perspective of our actual self seems contrary to the rationale for idealizing in the first place. I consider two solutions to the problem, one democratic, the other modelled on trusteeship. In the end, I argue, our actual self has complete freedom to choose the ideal self it grants rational authority. In the final part of the paper, I present my positive proposal as a solution to the underdetermination problem confronting the idealizing subjectivist and then argue that, so understood, this account vindicates a tidied-up version of how some reflective people already do deliberate in their everyday lives. This, in turn, suggests that a decision-procedure closely connected to the account is both possible (because actual) and attractive.

**Keywords** Indeterminacy · Underdetermination · Idealization · Rationality · Subjectivism

---

✉ Alex Horne  
aeh94@cam.ac.uk

<sup>1</sup> University of Cambridge, Cambridge, UK

## 1 Introduction

This morning, I have options.<sup>1</sup> Salient among them: work on this paper, read a novel, go for a run or call my folks. All seem permissible, none seem obligatory. It is tempting to think there is no fact of the matter about what I should do this morning, all things considered. Each option, let's suppose, is adequately supported by reasons of roughly similar weight (Portmore, 2011, 2013).<sup>2</sup> But this indeterminacy is untroubling. In part, because it is utterly unsurprising and in part, because I have time to perform all four options today. Also comforting is that the order in which I perform them hardly seems to matter. But, of course, there are many other valuable things I could be doing with my time: volunteering at a soup kitchen, writing that novel, coaching a little-league team, making my family breakfast, marking, reading a colleague's paper, donating money to charity, spending money getting to the Galapagos Islands, saving money for the future, giving-up philosophy to become a doctor and so on. There are many valuable things one can do in any given moment and in life generally, but one cannot do them all. So, at a whole-life level of analysis too, it is tempting to think there is no unique, best option and so, no fact of the matter about what I should do, all things considered. There will be many comparably valuable options and there will be incommensurable options and there will be incomparable options (Chang, 2002). Moreover, since every life contains some 'recharging time' and since the distribution of recharge time can vary greatly over equally valuable lives, it turns out that on any given morning I have just as much reason to sit on my couch drinking beer and watching *Sharknado* as I do volunteering at Animal Rescue (Gert, 2018; Portmore, 2013).

For those made of stern stuff, this indeterminacy might be liberating. Myself, I tend towards the existential dread end of the scale. But that's not the point. The point is that underdetermination threatens our ability to make non-arbitrary decisions about what to do. Flipping a coin to decide between equally appealing boxes of cereal seems sensible (Holton, 2006; Ullmann-Margalit & Morgenbesser, 1977). Flipping a coin to decide between different lives seems insouciant. If, moreover, what I should do today is a function of what I should do this week and what I should do this week, of this month and this month, of this year and so on, then indeterminacy is rife. It is not just that the weight of our reasons underdetermines which of a manageable number of options we should choose. It is that:

At any point in time in your life, many different choice situations are open to you, each of which you have sufficient reason to be in relative to agential values. If there is no more reason for you to be in one rather than another, the trajectory of your life, made up of the choices you make in arbitrarily selected choice situations, is itself profoundly arbitrary... But this makes a travesty of the human condition. (Chang, 2021, p. 105)

<sup>1</sup> For comments on earlier drafts of this paper I am indebted to Richard Holton, Yarran Hominh, Nikolas Kirby, Claire Kirwin, Nikhil Krishnan, Paulina Sliwa and *Synthese's* anonymous referees, together with its editors.

<sup>2</sup> In fact, Joseph Raz gives significant weight to what he labels 'the basic belief', namely, that 'most of the time people have a variety of options such that it would accord with reason for them to choose any one of them and it would not be against reason to avoid any of them' (Raz, 1999, p. 100). Cf. (Raz, 2012).

Call this ‘the rational underdetermination problem’. The existing literature on the rational underdetermination problem often construes it as one resulting from the ubiquity of objective values. It is therefore sometimes argued that subjectivists need not be troubled by the underdetermination problem. But on closer examination, it turns out, they should. Or so I will argue. The task of the first half of this paper is explaining why. The task of the second half is finding a subjectivist solution to the rational underdetermination problem. The basic argument is as follows: if subjectivists do not idealize, then their view is implausible.<sup>3</sup> So, subjectivists should idealize. But idealization is notoriously tricky (Dorsey, 2017; Enoch, 2005; Sobel, 2016).<sup>4</sup> Crucially for the argument of this paper, it is often difficult to generate a single, unified evaluative perspective from which to assess an agent’s options (Rosati, 1995). Our actual evaluative perspective underdetermines possible idealizations of it. Worse still, inter-perspectival comparisons are mostly blocked by incommensurability. So, I will suggest, idealization leads to indeterminacy and, since there is no further evaluative perspective one can occupy from which the deliverances of these other perspectives can be assessed, we are stuck. Each of us has many possible ideal selves, they all disagree about what to do, and too many cooks spoil the broth. In response: these idealized options should be assessed from the only perspective we can, in fact, occupy, namely, that of our non-ideal, actual selves. Deciding what to do therefore becomes, in part, an exercise in deciding who to be. But one might now worry this just moves the arbitrariness bump in the rug. Privileging the perspective of our actual self seems contrary to the rationale for idealizing in the first place. I consider two solutions to the problem, one democratic, the other modelled on trusteeship. In the end, I argue, our actual self has complete freedom to choose the ideal self it grants rational authority. In the final part of the paper, I present my positive proposal as a solution to the underdetermination problem confronting the idealizing subjectivist and then argue that, so understood, this account vindicates a tidied-up version of how some reflective people already do deliberate in their everyday lives. This, in turn, suggests that a decision-procedure closely connected to the account is both possible (because actual) and attractive.

## 2 Underdetermination by reasons

Not everyone is as squeamish about rational underdetermination as I am. Some even think it is a pro of objectivist theories that they generate significant normative wiggle-room. Joshua Gert provides an excellent summary of how this issue plays out between objectivism and subjectivism:

<sup>3</sup> Not least, one might worry, because the resultant overgeneration of reasons entails indeterminacy regarding what to do.

<sup>4</sup> For example, does idealization involve the actual agent in counterfactual circumstances of full information; an idealized deliberating agent in the actual circumstances of the present agent; or an idealized advisor considering the options of the actual agent? If idealization involves ideal deliberation, what exactly does that involve? Coherence constraints will obviously have a role to play, but does idealization also involve the revision of an agent’s current beliefs and attitudes in light of new beliefs and attitudes generated by imaginative acquaintance plus closure?

It is somewhat surprising that objectivism should go hand-in-hand with the idea that reasons—and the values that give rise to them—do not determine how we ought, rationally, to act. After all, objectivity seems intuitively to be associated with determinacy. Perhaps equally surprisingly, the converse of the link between objectivity and rational underdetermination also seems to hold. That is, if one takes reasons to be subjectively determined by one's contingent desires, it then turns out to be natural—not to say obligatory—to say that we have *fewer* rational options than the objectivist countenances. Why? Because it is more plausible that one's desires are characterizable in terms of a single, commensurable sort of strength than it is that objective values are. And if one's desires are characterizable in this way, and if one's reasons are a function of these desires, then it is a natural consequence that one's reasons also will have unique and commensurable strength values. To the degree, then, that one thinks that we do in fact have a great deal of rational latitude in our everyday choices, one has reason to reject subjective, desire-based theories of reasons, and to accept a more objective, desire-independent view. (Gert, 2018, pp. 459-460)

Of course, some subjectivists are likely to draw a different conclusion, namely, that it is a significant advantage of their theory that there are often determinate answers to questions about what one should do, all things considered, and that these answers are not alienating or arbitrary from the perspective of the agent subject to them. But, I will argue in the next section, the appearance of increased commensurability to which Gert alludes is misleading. As a result, subjectivists must also confront the rational underdetermination problem. Either that, or revel in rational latitude comparable to that enjoyed by their objectivist colleagues down the hall.

### 3 Subjectivism: why idealize?<sup>5</sup>

Idealization is central to two related but distinct classes of subjectivist theory, each with multiple variants. The first are full information accounts of well-being. The second are full-information accounts of reasons for action (Sobel, 2016, p. 115).<sup>6</sup> In each case, the basic rationale for idealization is that our uninformed desires can lead us astray (Heathwood, 2005; Hubin, 1996; Murphy, 1999). In the context of idealizing subjectivist accounts of reasons for action, Bernard Williams' celebrated example is that of my mistakenly believing a glass contains gin when, in fact, it contains petrol. I desire to drink it, but it seems wrong to say I therefore have a reason to drink it. If I knew it contained petrol, I would no longer desire to drink it. It is the desires of my better-informed self that determine my reasons for action (Williams, 1981). A similar rationale motivates idealization in full information accounts of well-being. Perhaps what I most desire right now is to be a famous athlete: the excitement, the money, the adoration! But, unbeknownst to me, I would in fact find the life of endless training required boring and shallow. Worse still, I would find the public scrutiny of my every

<sup>5</sup> See (Enoch, 2005).

<sup>6</sup> Cf. (Brandt, 1979), (Darwall, 1983), (Dorsey, 2017), (Gibbard, 1990), Lewis, D. (1989), and (Railton, 1986a).

moment and the pressure to perform on game-day incredibly anxiety-inducing. It is not what I think I desire, but what I would desire were I properly acquainted with it, that determines the good for me. In cases like the desire to be a famous athlete, I am mistaken about what it would actually be like to *be* a famous athlete so what I now desire is a chimera. I really do desire the life of an athlete *as I imagine it*, but not the life of an athlete as it really would be.

In both cases just discussed, idealization plays what might loosely be described as a ‘negative’ or ‘defeater’ role. I desire something because I am mistaken about its true nature. Were I properly acquainted with the true nature of the object of my desire, I would no longer desire it.<sup>7</sup> Here, the non-idealized agent’s misapprehension is highlighted by the idealized agent’s choice and defeated by it. That is, the ideal agent’s preference acts as a *disabler*. In this guise, idealization does indeed look like it might lead to greater determinacy. Fanciful, poorly thought-out options are trimmed from my deliberative purview: I no longer have reasons to be a dinosaur or an eighteenth century pirate, astronaut, firefighter or wizard, though all of these have been aspirations of mine at some stage of my life. All that remains, let’s suppose, are more “grown-up” desires for health, happiness, intellectual stimulation, enriching experiences, meaningful relationships, perhaps doing a bit of good and so on. Four observations are in order. First, at this level of desiderative generality and even allowing for my personal quirks, dispositions, tastes and the like, there are a bewildering number of utterly different lives I might lead that would contain satisfaction of these various desires in different ratios but with roughly the same degree of overall preference satisfaction. Second, how the details are to be filled in is precisely where one might have hoped deliberation on what to do might help, but it is at precisely this point that overall rough parity between total welfare across the candidate lives deprives one of principled means of distinguishing between them and flipping a coin seems utterly unsatisfactory for so momentous a decision. Third observation: at this point, idealization plays what I will label its ‘positive’ or option-enhancing role, and so at this point increases, rather than trims, indeterminacy about what to do. That is, what would otherwise be unreasonable options are enabled by the choice of an idealized agent that makes them reasonable. Here, the ideal agent plays the role of *enabler*. Suppose that I have never tried a kiwi-fruit, am unaware they exist and so presently have no desire to try kiwi-fruit, but that, were I to try one, I would love it and come to desire it as much as any fruit with which I am presently acquainted. Most proponents of idealizing subjectivism maintain, very plausibly, that I therefore have some reason to try kiwi-fruit and that trying kiwi-fruit would be good for me. It would be unduly impoverishing to limit our good to options of which we happen to be aware. But now the extent to which subjectivism provides a bulwark against indeterminacy is seriously in doubt. If I now have reason to perform each option that I would desire to perform were I aware of how the experience would affect me, and the aim is then to generate some unique combination of options that would determine what I should do with my life, the chances of success—even of generating a manageable range—are slim indeed. The fourth observation is that, insofar as the rational underdetermination

---

<sup>7</sup> Of course, this is perfectly consistent with my continuing to desire a glass of gin or the life-of-an-athlete-as-I-imagined-it.

problem is concerned, it now looks like subjectivists and objectivists are presented with comparable deliberative scenarios. Perhaps I presently desire to steep myself in a world of historicism, whimsy, and camaraderie, but am unaware that joining my local Morris dancing troupe is an optimal way of achieving this, though I am aware of the troupe's existence. Or perhaps I do not have any present desires for historicism, whimsy, and camaraderie or Morris dancing but, were I to take-up Morris dancing I would love it for the historicism, whimsy, and camaraderie it provides, filling a hole I did not know was there. Once we are contemplating what we *would* desire, were we acquainted with it, there is a significant degree of overlap between the options subjectivists and objectivists will find they have. After all, there is a significant degree of overlap between what objectivists and subjectivists find *valuable*.

#### 4 Idealization and underdetermination

If the argument so far is on track, then *pace* Gert's suggestion, there is reason to think there may be little difference between the degree of indeterminacy facing subjectivists and objectivists on the whole. In fact, we would expect this to vary with the preference structures and desiderative profiles of different agents. In this section, I argue that subjectivists are actually confronted with *more* indeterminacy than objectivists. The basic idea is that subjectivist idealization is not subject to deontic constraints that rule out otherwise internally coherent but intuitively unappealing combinations of desires. This problem is then compounded by incommensurability between the perspectives generated by idealization.

To illustrate the increased indeterminacy facing idealizing subjectivists post-idealization, consider the case of an agent we'll call 'Greenie'. Greenie desires health, happiness, intellectual stimulation, enriching experiences, meaningful relationships and so on, like many agents (Railton, 1986b). Additionally, Greenie has a very strong desire to protect the environment. Given her desiderative profile, both objectivists and subjectivists will, I assume, agree that various lives could be equally good for Greenie: perhaps there is nothing to choose between Greenie's becoming a conservationist living in Rangat, a marine biologist living in Port-aux-Français, a climate scientist living in New York, and a green-tech investment banker living in Brussels. But we might expect objectivists to balk at other possibilities equally a function of her desiderative profile, say, becoming a violent eco-terrorist or leader of a euphoric enviro-anti-natalist-death-cult. Presumably, this is because there are objective reasons not to risk or intend to damage others' property, kill people, encourage others to kill themselves in the name of environmentally-motivated population control and so on. This is a variation on the theme of what Sharon Street calls 'ideally coherent eccentrics' (Street, 2009). Since, we may suppose, it is possible that Greenie would be happy as an eco-terrorist, would find it a meaningful existence full of fulfilling relationships with other eco-terrorists, enjoy the intellectual stimulation of planning eco-terrorist operations, would find carrying out these operations full of enriching experiences and be filled with an overwhelming sense of purpose and so on, it actually looks like the idealizing subjectivist has more options before her than the objectivist. Sure, we might think, Greenie is not right now contemplating blowing up coal-fired power-stations,

but that is only because she has not realised that this too is well-supported by her reasons, given her desires. Blowing up coal-fired power stations is, one might argue, like kiwi-fruit or Morris-dancing. Don't knock it, 'til you've tried it (in the fully-informed imagination of your ideal self).

At this point, you may be thinking that no one self—even one with full information and idealized deliberative and imaginative capacities—could simultaneously entertain the prospect of being a mild-mannered, family orientated marine-biologist and a cut-throat eco-terrorist while holding onto a sense of how to compare them. I suspect there is some truth to this conjecture, and further reflection on it suggests a further potential indeterminacy problem for the idealizing subjectivist, by way of a dilemma. On the one hand, let's suppose that there is just one ideal self that can indeed contemplate and find equally attractive becoming one of many different selves as a result of pursuing an optimal, idealized option. In that case, if the actual deliberating agent should do what its ideal counterpart would advise, say, then the actual deliberating agent has equally weighted reasons to become any one of its ideal counterparts: biologist, banker, conservationist, terrorist, climate scientist, cultist and no way to choose between them. On the other hand, suppose that it is not possible to integrate these various possible idealizations into a single perspective. In this case, the actual deliberating agent has multiple ideal selves each of which has a different idea of what would be best.<sup>8</sup> Some of the moderate ideal selves may well be able to understand the appeal of other moderate ideal selves: the marine biologist and the conservationist will, we might think, probably get along. But other selves will find one another utterly mystifying and repulsive, in part precisely due to their shared starting point. The eco-terrorist will be disgusted by the self-serving comfort and meagre contribution of the conservationist, while the conservationist will, one assumes, be horrified by the violence and zealotry of the eco-terrorist. The actual deliberating agent, contemplating which course of action to pursue, is at this point confronted not just with equally weighted options and no principled decision procedure, but with incommensurable conceptions of the agent's own good, despite each conception resulting from idealization of her actual evaluative perspective. In short, the actual agent's desiderative profile either underdetermines the options available to its ideal self or underdetermines the agent's ideal selves. Either way, indeterminacy looms.

In fact, the situation facing the idealizing subjectivist may be trickier still. Throughout the discussion of Greenie, I have granted, *arguendo*, a tacit preservative assumption. This preservative assumption is that an idealization of an agent's actual evaluative perspective will be maximally preservative of her existing desires and values, integrating *all* of them, to the extent that it is possible to do so, into a unified evaluative perspective. So, if an agent desires and values health, happiness, intellectual stimulation, enriching experiences, meaningful relationships and protecting the environment, then each of her ideal selves will also desire and value these things and the lives they choose will reflect this. Different equally rational selves—and, therefore, options—will do better and worse *vis-à-vis* different desires, we might think, but

---

<sup>8</sup> Cf. (Walden, 2015)'s reflections on open-ended constructivism with (Sampson, 2021), who takes something like this argument to present a direct challenge to reasons internalism. But given that rational underdetermination is a pervasive problem for externalists about reasons as well, it is difficult to see why this should be a *special* problem for internalists.



what matters, let's suppose, is overall satisfaction. Perhaps the green-tech investor has much higher disposable income facilitating better health treatment and many more enriching experiences than the conservationist, but the conservationist enjoys much richer personal relationships and so on. But the conservative assumption looks like it requires independent justification. This is not the Millian point that it is better to be less than satisfied Socrates than a very satisfied pig, rather, it is not a *given* that the very same rationale that permits differential levels of satisfaction of each desire, provided overall satisfaction is equal, might not permit compensation for the loss of a desire at the ideal level (or its dissatisfaction at the non-ideal level), with a massive increase in the satisfaction of the others desires and values an agent has (Mill, 2009). Perhaps becoming a hugely successful green-tech investor will instead have deleterious consequences for the investor's health in the long run, but this is more than compensated for by corresponding gains in the richness of the experiences and success in protecting the environment thereby achieved. But it seems odd to think this not a rational option for Greenie, because Greenie also values her health. Given Greenie's other values, this seems a permissible trade-off. But at this point, the indeterminacy worry presses further still. After all, if candidate options need not satisfy each and every desire and value and agent has—and it is implausible to think they must, given how many desires each of us has, and the difficulty of holding them all together coherently, let alone realizing them in our actual lives—then there are many, many more options on the table than we might first have thought, while holding onto the tacit assumption that we preserve as many existing desires or values as possible. For example, if Greenie desires to both protect the environment *and* alleviate the condition of the working-poor in her hometown *and* to care for her elderly parents, it is going to look much more difficult to integrate satisfaction of every desire into every rationally justifiable option or ideal self. For the idealizing subjectivist, Sartre's case of the torn dutiful-son-resistance-fighter generalises across many more quotidian cases.

To re-emphasise, I am not saying this is a problem for idealizing subjectivism—it may be that this is just the way the cookie crumbles, normatively speaking. But I am saying it is a mistake to think subjectivists are *prima facie* any “better” off when it comes to facing determinacy than objectivists, in fact, things may be significantly “worse”. I employ scare-quotes here because, of course, if one thinks this really is how things are or, like Gert, thinks that delivering a highly permissive account of rationality is attractive, then subjectivists are actually *better* off in this regard than objectivists. So, if this is part of the motivation for being an objectivist, it might be worth switching sides.

#### 4.1 Tie-breakers?

But perhaps I have been uncharitable—or, given the final point of the previous section, *overly* charitable—in my assessment of what idealization achieves. Specifically, the subjectivist could argue that for an idealization of a given agent's actual evaluative perspective to count as *genuinely* ideal it must meet certain further conditions and that meeting those conditions will very likely exclude all possible idealizations bar one.



In the previous section I prefaced an objection to this strategy. I argued that breaking a deadlock between (equally) ideal selves—or claiming that one was *more* ideal than all others—would require appealing to some external, objective (deontic or otherwise) value and that this was *prima facie* inconsistent with subjectivist theories of well-being and reasons. There are, however, two ways an idealizing subjectivist might try to rebut this *prima facie* inconsistency. The first is to deny that some values—theoretical virtues like elegance, conservatism and simplicity—are external and objective in the relevant sense. These values, the idealizing subjectivist might argue, are certainly external and objective in one sense. But in another sense, they are not external to but rather constitutive of the very normative practices idealizing subjectivist theories are trying to explain. Appeals to such values are necessary in order to even understand what it means for one evaluative standpoint to be ideal or for what counts as idealizing an actual evaluative perspective instead of, for example, arbitrarily substituting some internally coherent but alien and un-related evaluative perspective for a person's actual evaluative perspective. These values are therefore not external and objective in the sense to which the idealizing subjectivist objects.

But although I think this account of these purportedly external and objective theoretical values might allow idealizing subjectivists to avoid the challenge that their justification for appealing to them is objectionably ad hoc, I think this comes at the cost of the relevant values' potential utility in breaking the deadlock between different possible ideal selves. The reason characterising theoretical values in this way will undermine their use in prioritising one ideal self over another is as follows. If the values in question are constitutive of the normative practices of which the idealizing subjectivist is trying to provide an account and so are, in turn, the very norms by which idealization is to be understood, then they cannot be used to decide which ideal self should take priority over the others. By definition, each ideal self meets the norms constituted by these values equally, otherwise the self in question would not be ideal. For the idealizing subjectivist's characterisation of these objectivist-looking values as constitutive of the enterprise to work, it must be the case that, for example, ideal coherence is much more than just mere consistency. Ideal coherence will then be constituted not just by perfect consistency plus full information but also by, for example, maximal elegance.

The idealizing subjectivist might then have hoped to use these values to measure degrees of idealness but, to the extent that the values can be put to that use, they simply rule out counting some purported idealizations as genuinely ideal in the most robust sense. But this does not provide a way of prioritising one ideally coherent self over other equally coherent and ideal selves, which is what they require to avoid the underdetermination problem. Rather, this is just an explanation of the way in which we might rank different possible selves against a metric of how close to ideal they really are, namely, by considering the extent to which they meet the relevant theoretical virtues. In order for this to help the idealizing subjectivist, they must provide a further argument to the conclusion that in most or all cases, there will be one self that is "more ideal" than all the others. The mere fact that idealization is a rich and complicated process is not sufficient to render this claim plausible. If I am right about this, then an appeal to theoretical virtues will not break the deadlock between selves, even if the idealizing subjectivist can consistently appeal to such values in this context.

## 5 Conservatism

But there is a second route the idealizing subjectivist might take to rebut the *prima facie* inconsistency between subjectivism and appeal to such apparently external values in order to break deadlocks between ideal selves. The second route is to argue that the value of *conservatism* is primary and is capable of generating a unique ideal self. This is analogous to its use as a tie-breaker in cases of underdetermination of scientific theory. Some argue that where a choice must be made between two equally coherent theories, we should, or at least do, prefer the one that requires the least alteration of our existing belief set.<sup>9</sup> So, we might claim that one ideally coherent self has greater normative authority over us than any other because it preserves the greatest number of the values, preferences and dispositions that were constitutive of our starting actual evaluative standpoint. In other words, the idealizing subjectivist might claim, we should do what our ideal self would want us to do, where our ideal self is the self that is both ideally coherent and most similar to our actual self.

Before assessing the extent to which this appealing suggestion can be squared with the idealizing subjectivist's other commitments, I note that, like the other theoretical virtues, conservatism is open to the challenge identified above. That is, either conservatism is an external value of the sort to which the idealizing subjectivist cannot legitimately appeal or it is constitutive of the relevant normative practice: one of the values by reference to which idealization should be understood. If the latter, the idealizing subjectivist must argue for the claim that this value will (a) generate a unique ideal self and (b) that this self always trumps other potentially ideal selves that are generated by placing less weight on the value of conservatism and more on, say, elegance or creativity. In any event, there are reasons to doubt that even were conservatism granted primacy as an organising value, it would generate a unique result. For example, how are we to understand which starting preferences, dispositions and values count for the purposes of this measurement? By number, number of entailments, strength, duration or some other metric? As a matter of fact, it seems likely that for many of us there will be multiple equally conservative ways of idealizing our actual evaluative standpoint. If that is true, then conservatism will not do the trick.

I doubt that this problem can be overcome. But even if it could, I believe the conservatism proposal is mistaken, albeit tempting. It is tempting for at least two reasons. The first reason is that giving the value of conservatism priority generates an intuition that there will be more values left at the end of the idealization process and that the ideal self will be correspondingly more substantial. If what I have been saying about idealization is correct, then ideal coherence can come cheaply by simply deleting and adding values from our actual evaluative standpoint. A concern one might therefore have is that many of our ideal selves will be extremely sparse: single-minded and devoted to a cause. Sure, they are ideally coherent in one sense, but there is not much to them. One might think, then, that the ideal self that conserves the greatest number of our existing values will be a much richer and more complex self and is, for that reason, preferable.

<sup>9</sup> See (Quine, 1980) and (Quine, 1976). Quine argues that our conservative disposition is a matter of fundamental psychology, though others argue that conservatism is also a theoretical virtue: Cf. (Sklar, 1975).

But there are three problems with this thought. The first problem is that it is not clear that conservatism will generate richer or more substantial selves. Idealization means that there are as many ideal selves as there are consistent starting sets of desires and values in a person's actual evaluative perspective (where a set can, of course, be a singleton). But once that starting point is adopted, the ideal self is constructed from not just the starting value(s) but also those values that will follow from and cohere with them. As Smith observes when building on Williams' account of idealization:

Williams is right, I think, that deliberation can both produce new and destroy old underived desires. But he is wrong that the only, or even the most important, way in which this happens is via the exercise of imagination. For by far the most important way in which we create new and destroy old underived desires when we deliberate is by trying to find out whether our desires are *systematically justifiable*. (Smith, 1994, pp. 158-159)<sup>10</sup>

It is for this reason that idealizing subjectivists can argue that a person's ideal self would advise them to perform some action that is only related to his or her actual evaluative perspective tangentially, in the sense that the relevant desires are consequences of having the other desires they do. This means that a great many of our ideal selves will be just as rich and complicated as the ideal self that follows from idealization plus conservatism.

The second problem is that it now looks like we are trying to justify an appeal to conservatism in terms of another value, one that looks decidedly objective and external to the self in question. That value we might call 'complexity' or 'substantiality' or 'richness'. We now need some further explanation of, for example, how we should weigh the single-minded pursuit of some ideal against a complicated life full of compromise, together with an explanation of why we should prioritise the latter. It is difficult to see how this explanation can proceed without appealing to further external values in deciding between the different ideal selves.

The final problem is that conservatism is *not* just a proxy for richness. By placing conservatism at the centre of our account of idealization, of course, we prioritise our actual self at any given moment, but what reason do we have to do that?

The obvious answer to this question is that our actual self is, as a matter of fact, important to us. This is the second reason that the idea that our ideal self is the ideally coherent self closest to our real self is tempting. We care about our actual selves in a special kind of way. Our concerns, values, preferences and dispositions at this point in time are important to us because they are *ours*. This is one of the reasons subjectivism is attractive in the first place. Objectivism ignores our special concerns and plans and gives priority to alien rules and requirements in their stead. One reason, then, to give the ideal self that is closest to our actual self priority is that it is the *least* alien, "external" ideal self. Correspondingly, its authority seems both more legitimate and less demanding than all our other ideal selves. Not only that, but we also might have a better chance of grasping what our ideal selves want us to do if they are not too remote from our actual evaluative standpoint. If this were true, then perhaps this might justify

<sup>10</sup> Of course, given the fact of underdetermination, I deny that trying to find out whether our desires are systematically justifiable entails conservative, Rawlsian reflective equilibrium.

giving the closest ideal self priority and, if we can give one ideal self priority, then underdetermination is no longer a problem.

Unfortunately, I do not think this will work, for three reasons. The first reason comprises the familiar set of problems associated with indexing what you should do to your present desires and beliefs. The rationale for idealizing in the first place is that we are often mistaken about what we think we desire, know and value. For some people (either very lucky or very wise or both) it may be that some or all of their ideal selves are very close to their actual self. But for most of us, it seems very likely that most of our ideal selves will look very different from who we are now. Closeness to our actual self might be intuitively attractive in some cases and unattractive in others—how far each idealization takes us from our current evaluative standpoint will depend on the details of each person's psychology and place in the world—but there does not seem to be a non-circular way of justifying the priority of our actual self, at least not without appeal to some external value. Privileging those selves that are closest to the starting point would only make sense if the starting point had some normative significance over and above its being a set of psychological facts about us, but, by hypothesis, it does not.

But even supposing I was mistaken about that, giving priority to an ideal self because it is closer to an actual self appears to involve the following error: it supposes that some of your ideal selves are more you than others. But that is a mistake. Each of our ideal selves is us, but better, in the sense of having full information and reasoning infallibly. It is no objection to idealizing subjectivism generally that some people's ideal selves look very different from their actual self. On the contrary, that fact demonstrates the power of idealization in making sense of our normative practices. Some people lack more information, make more mistakes and start with more incoherent sets of values, preferences and dispositions than others. It is therefore a mistake to think that, given underdetermination, one ideal self enjoys lexical priority over others as a result of sharing surface similarities with an actual self at a given moment in time. Each ideal self has an equal claim on us in virtue of being a fully-informed and error-free version of ourselves. That is why underdetermination is a big problem for the subjectivist too.

If the argument so far is correct, then no ideal self enjoys priority over any other just in virtue of its proximity to our actual self. While I believe this result exhibits deep consonance with the subjectivist project, the subjectivist's underdetermination problem remains. In the final sections of the paper, I sketch two potential solutions to it. The first is democratic and the second is modelled on trusteeship. I argue that potential technical problems with the democratic solution are soluble. But I also argue that, once solved, the solution looks difficult to square with a key reason for becoming a subjectivist in the first place, namely, the sense that there should be some close connection between an agent's actual desires and their reasons, such that the latter are capable of *motivating* actual agents in the right circumstances. This is not intended as a decisive objection to the democratic solution. But it does, in my view, give us a good reason to prefer a solution that looks more amenable to this characteristic subjectivist commitment. The trusteeship solution, on the other hand, satisfies this motivational desideratum quite well. In my view, this gives us a reason to prefer the trusteeship solution.

## 6 The democratic solution

I will label the first potential solution to the subjectivist's underdetermination problem 'the democratic solution', the bare bones of which are as follows. Imagine that each of our ideal selves gets one vote regarding what we should do on any given occasion. While they all disagree about what to do initially, let's suppose that through a yet-to-be-specified procedure of judgment aggregation they can arrive at a decision about what we should do. The democratic solution to the underdetermination problem maintains that we should do whatever the outcome of that hypothetical procedure says we should do.<sup>11</sup> That is, we should do what our ideal *selves* want us to do, where they are deliberating together as a *demos*.

There is much to like about the democratic solution. But, of course, the devil will be in the details. Judgment aggregation is not a simple matter, as Condorcet's paradox shows (Condorcet, 1785).<sup>12</sup> But, as a first pass, we should at least *start* simple. The simplest voting procedure applicable to our scenario is Plurality Rule, also known as 'First Past the Post'. In any deliberative scenario, each idea self would vote for the candidate action they believe the actual self should perform. The action the actual self ought to perform is the action that receives the most votes. So, if you are deciding between granola and blueberry pancakes for breakfast and more selves want granola for breakfast than blueberry pancakes, granola is what you should have. This doesn't mean you should always have granola, of course. It may be that the *demos* feels differently about the relative merits of granola and blueberry pancakes on weekends, for example. It just means granola wins on this occasion. But, of course, there are also some well-known problems with Plurality Rule, including that it sometimes picks a Condorcet loser as the overall winner. Some voting methods seek to address this problem by introducing quotas, for example, requiring a Strict or Absolute Majority winner. But addressing Plurality Rule problems with quotas does not guarantee a winner (and of course may make it less likely in some scenarios). So, it looks like there will still be some decisions where the underdetermination problem remains. Of course, the real problem here is that there are more breakfast foods available than just pancakes and granola and, if we raise the stakes to deciding on *careers* or the like, the possibility of deadlocks looks very real. In short, the existence of multiple ideal selves is predicated on their having conflicting views about what is best for the actual self, so *by hypothesis*, in many important deliberative scenarios, Plurality Rule will produce no result. And this is precisely what we should expect, if the argument of the first part of this paper is correct.

But there is another reason to reject Plurality Rule, which is that it is too coarse-grained. We will get a much more nuanced picture of the general will of the ideal selves if they are required to *rank* the candidate actions for the actual self to perform. Of course, in non-ideal voting scenarios, a significant drawback to ranking systems of voting is that they are very demanding: they require more time, effort and information than other means of decision in order to accurately reflect the preferences of the voters.

<sup>11</sup> Thanks to an anonymous referee for suggesting this possibility. I note that (Southwood, 2019)'s discussion of 'contractualism for us as we are' might be re-purposed to help with some of the problems outlined here.

<sup>12</sup> For a helpful overview of the judgment aggregation literature see: (List & Puppe, 2009).

But, since the ideal selves are ideal and, we can assume, have all the information required to make their rankings, know their own preferences immediately, and can compute all of this instantaneously, we need not worry about these drawbacks here. The best-known ranking system of voting is the Borda Count, in which, if there are  $n$  candidates, the voter assigns  $n - 1$  points to their preferred candidate,  $n - 2$ , to the second, and so on (Borda, 1784). So, applying this to our scenario, the Borda Score of a candidate action is the total points assigned by all ideal selves to that action and the action the actual self should perform is the candidate action with the most points. Will selecting this voting procedure—or another like it—ensure that there is always a unique answer to the question what the actual self should do? Let me start with an optimistic answer with an interesting result, before moving to pessimism as a reason to at least consider an alternative solution.

The optimist answers as follows. Yes, a procedure like Borda count; or a multi-stage voting method that eliminates least preferred options; or a voting procedure that allows ideal selves to provide more information about the strength of their preferences by allocating points in a different way between candidates; or some other appropriate procedure, will *always* provide a determinate answer.<sup>13</sup> We just have to find the right procedure. Supposing there is one to be found, can we say anything more about the sort of answers it might give to questions concerning what we ought to do? One hunch is that we might expect some tendency towards the mean. That is, the action the ideal selves will want the actual self to perform will be, on balance, the action one of the “middling” or “average” or “less radical” selves would have wanted. So, in the case of Greenie discussed earlier, we might expect the terrorist and cult-leader options to be ruled out, while the conservationist, marine biologist, climate scientist and green-tech investment banker look much more likely. In fact, *if* these professions could be ranked along a *single* spectrum and *if* a voting method were selected that satisfied the Condorcet Criterion (i.e., a Condorcet winner is always chosen if one exists), then Median Voter Theorem tells us that that the successful candidate will be that closest to the median voter (Black, 1948). But, of course, these are big ‘ifs’. For example, it is not obvious that Greenie’s candidate professions could be ranked along a spectrum of something like ‘more to less extreme’, even if we have the sense that terrorist and cult-leader are more extreme than the others. And even if they could, why is ‘extremity’ the relevant spectrum along which the candidate careers should be arranged? But perhaps we do not need the technical assurance of the Median Voter Theorem. Instead, there seems an intuitive sense of preferential “cross-cancelling” that suggests that the more radical options will get less support overall than those less radical, and so, the optimist says, there is some intuitive support for the idea that the demos of ideal selves will select the sensible centre candidate action from among the various possibilities. If that is true—and *if* it is possible to identify the sensible centre candidate—then not only does the democratic solution solve the subjectivist’s underdetermination problem, it does so in a way that might give us, for any given agent, some sense of which action

---

<sup>13</sup> These options are not intended as exhaustive, of course. There are many possible voting methods and the literature on their relative merits is vast, technical and well beyond the scope of this paper. Perhaps, if the democratic solution is independently plausible, it can wait for the best solution to these problems to be provided in that literature.

they ought to perform, provided we have an antecedent sense of who their possible ideal selves are.

The pessimist denies that any judgment aggregation procedure will always deliver a single winner. Their pessimism may stem from the vast literature on Arrow's Theorem (Arrow, 1963) or it may follow more straightforwardly from a sense that the different options a person has about what to do with their life are not the sort of thing that can be ranked: there are simply too many incommensurabilities and incomparabilities and ties in play to hold out any hope of always getting a determinate answer from a vote. For the pessimist, the problem with the democratic procedure is that it looks like it might give determinate answers to questions that don't really matter—like what to have for breakfast—and then leave us high and dry when making the decisions that really count. So, the pessimist worries, the democratic solution is at best a partial solution to the subjectivist's underdetermination problem. As it turns out, the questions that really matter do not have positive answers, though some courses of action will be ruled out.

But while I think the pessimist is right to worry that a straightforward judgment aggregation procedure is unlikely to generate determinate answers all the time, I also suspect that the rich resources of social choice theory will ultimately be up to the task. That is, I am fairly confident that some procedure could be found that solves all the technical issues with the democratic solution to the subjectivist's underdetermination problem. But even supposing that it can—and in part precisely *because* resolving technical issues in social choice theory is complex—I think there is a much more pressing, principled objection to the democratic solution.

The principled objection is that the degree of complexity and abstraction the democratic solution involves cuts actual, fallible, cognitively limited agents off from their reasons in two ways: *motivationally* and *practically*. Part of the initial attraction of subjectivism was supposed to be preserving a close connection between our actual desires, motivations, values and concerns and our *reasons*, such that those reasons were somewhat *tractable* for actual agents and, when apprehended, actually *motivated* us to perform the action they support.

Now, of course, the metaphysics of reasons need not straightforwardly deliver a decision-procedure, but one might hope it doesn't make deciding correctly look beyond the reach of most actual people either. But if the correct answer to what one should do at any given moment requires instantaneous, complicated social-choice-theoretic computations concerning preference aggregation across large numbers of ideal selves, then most people are—as a *practical matter of epistemic access*—cut-off from their normative reasons most of the time. That is, most of us will never actually know what we're supposed to do at all, let alone in time to make a decision on the spot.

Worse still—and even supposing I were mistaken concerning practical epistemic access—the democratic solution risks severing a crucial connection between reasons and motivation. That is, supposing one could ascertain the decision handed down by one's demos of ideal selves, there doesn't seem to be anything *irrational* about the agent who is then utterly *unmotivated* to act on the reason they're given. In part, this will be due to the bewildering technicality of the metaphysics of their reasons. In part, it will be because the agent will know that while one action may ultimately have won the day, the reasons each ideal self may have for preferring that action may be



completely different from and perhaps even straightforwardly inconsistent with the reasons other ideal selves have for preferring that action. This seems to go some way to stripping their verdict of its authoritative motivational appeal, such that it no longer seems bizarre or irrational to fail to be motivated by it, given one's other motivations. But preserving this sort of close motivational connection was one of the main reasons to become a subjectivist in the first place. So, if this worry is on the right track, the problem with the democratic solution to the subjectivist's underdetermination problem is that it deprives subjectivism of much of its initial appeal *and* is unlikely to be found congenial by many subjectivists. This is not a decisive objection to the democratic solution. It may be that this is just the way things are, normatively speaking. But I think it is troubling enough to hunt for another solution that scores better on these practical and motivational metrics. And I think the trusteeship solution does just that.

Let me briefly recap. The democratic solution to the subjectivist's underdetermination problem faces technical problems concerning the possibility of ties. That is because if there are often ties, then underdetermination and indeterminacy permeate the ideal level of decision-making too. But it is likely these technical problems can be solved. There is, however, a more pressing objection, namely, that the democratic solution cuts actual agents off from their reasons as a practical matter of epistemic access and by severing the close connection between reasons and motivation that is at the heart of subjectivism. This is not a decisive objection to the democratic solution, but it does give a subjectivist who wants to preserve the close connection between reasons, motivation, and actual decision-making a reason to look for an alternative solution.

## 7 Choosing yourself

Let me therefore suggest an alternative solution to the underdetermination problem. I will label it 'the trusteeship solution'. I will begin by explaining it in the abstract as a solution to the underdetermination problem confronting the idealizing subjectivist and then argue that, so understood, this account vindicates a tidied-up version of how some reflective people already do deliberate in their everyday lives. This, in turn, suggests that a decision-procedure closely connected to the account is both possible (because actual) and attractive.

The basic idea, in the abstract, is as follows. *We should do what the ideal self we would choose to be, were we aware of their existence, would want us to do.* Let me explain. Suppose for a moment—*per impossible*—that each actual deliberating agent were aware of every one of their possible ideal selves: a virtual smorgasbord of candidates for who to become as the best version of themselves. Now, since no ideal self enjoys *automatic* priority over any other possible ideal self, the actual self is not subject to the automatic rational authority of any one of its ideal selves either. Each is a candidate for the allegiance of the actual deliberating agent, who can decide which ideal version of themselves to entrust with rational authority over their life. Each ideal self represents an ideally coherent, fully-informed version of them, with a clear plan for what they should do, integrated into an evaluative standpoint that determines what is valuable in life and how to live well.

Why is this second solution to the underdetermination problem modelled on *trusteeship*? In brief, the idea is that the relationship between actual and ideal agent is akin to a form of non-delegatory representative government.<sup>14</sup> It is not a form of *delegation* because the ideal self does not simply follow the present preferences of the actual self. They are not a mere mouthpiece for the actual self because, by hypothesis, many of the preferences of the actual self are inconsistent with the ideal self, otherwise idealization would be otiose. Instead, the trustee ideal self is bound by fiduciary duty to govern on the basis of its best understanding of the interests of its beneficiary, the actual self. Like a trustee, the ideal self has unfettered discretion to select both the ends to be pursued and means appropriate to pursuing them.

These brief remarks also suggest an answer to the following objection to the trusteeship solution. If it is up to us to decide which ideal self determines our reasons, is it even *possible* for us to make a mistake by switching between ideal selves post-decision? After all, if we commit to one ideal self, but then behave perfectly in accordance with the demands of another, we're still acting in perfect accord with the recommendations of an ideal self, so how could we possibly be making a rational mistake? But if we are not making a rational mistake, then doesn't this show that the trusteeship solution to the underdetermination problem fails?<sup>15</sup>

Responding to this objection requires thinking carefully about the diachronic relationship between synchronic ideal selves. Suppose, simplifying grossly, that at  $t_1$  you have a choice between two ideal selves,  $A_1$  and  $B_1$ , and that you choose  $A_1$ . Shortly thereafter, at  $t_2$ , you begin behaving in accordance with the recommendations of  $B_1$ . In what sense have you made a rational mistake, since  $B_1$  was among your ideal selves at  $t_1$ ? The answer is that  $B_1$  is no longer among your ideal selves, because it is no longer  $t_1$ . There *may* be an ideal self at  $t_2$  that resembles  $B_1$ , and for argument's sake, we will call this ideal self " $B_2$ ". What are the salient differences between  $B_1$  and  $B_2$ ? Well, unlike  $B_1$ ,  $B_2$  is an idealization of your actual evaluative standpoint at  $t_2$ , which includes your prior commitment to  $A_1$  and so on. Now, *if*—even allowing for that commitment, and the value you place on commitments (if you do),  $B_2$  is an ideal self—*then* no, you have not made a mistake in behaving in accordance with  $B_2$ 's recommendations, because that ideal self has "built-in" the breaking of rational commitment, the switching costs and so on. But it *would* be irrational to behave in accordance with the recommendations of  $B_1$ , because at  $t_2$ ,  $B_1$  is no longer among your ideal selves.

But perhaps the *really* troubling feature of the trusteeship solution is not the unfettered discretion of the trustee ideal self, but the *unfettered discretion of the actual self* to select their ideal self. It is to this worry I now turn. Recall that in the hypothetical decision scenario we have complete freedom to choose the ideal self we entrust with rational authority over ourselves on the basis of the intrinsic appeal of that ideal self,

<sup>14</sup> On the distinction between delegation and trusteeship in political representation see (Dovi, 2018).

<sup>15</sup> Thanks to an anonymous referee for pressing this objection. Of course, depending on one's view of the role of intentions in practical reasoning, it may be that it is straightforwardly irrational, because means-end incoherent, to intend to but then fail to take the steps necessary to ing. If this is true then the actual self that intends to commit one way but then behaves inconsistently with this commitment has made a rational mistake. Whether this is also therefore a normative mistake is a point of significant contention. Cf., for example, (Bratman, 1987), (Broome, 2001), and (Kolodny, 2011).

viewed from the actual evaluative perspective of the fallible, cognitively limited and less than fully informed and coherent agents we are.<sup>16</sup> This is because each candidate is “equally” ideal: our rational freedom is only constrained by the structural rationality that renders each potential self ideal, which, by hypothesis, it already is. But shouldn’t we ask why the actual self should have any say in the matter *at all*. Doesn’t granting the final say to the actual self undermine the solution altogether? If the motivation for idealization in the first place was that the actual self is poorly informed, irrational, weak-willed, and motivationally incoherent, why on earth should we grant this unimpressive specimen ultimate normative authority?<sup>17</sup> But although this objection touches on something deep and troubling about how best to justify assigning any value whatsoever to the desires and preferences rational agents just happen to have, I do not think, on closer inspection, it undermines the trusteeship solution to the *underdetermination problem*. This is because, I suggest, on the best understanding of the dialectic in this paper, the problem is: the underdetermination problem, given idealization; and *not*: why idealize, given the underdetermination problem.

The rationale for idealization in the first place is the irrationality and imperfect information of actual deliberating agents. Idealization is not generally directed towards solving the underdetermination problem, rather, it is directed towards solving the problem that, on the one hand, the subjectivist believes reasons and an agent’s welfare are grounded in facts about what they (would) desire, while on the other hand acknowledging that unformed and irrational desires are a poor basis for determining an agent’s reasons and welfare. Idealization is supposed to cure the deficiencies in our subjective motivational set to avoid gin-petrol cases and the like. An assumption shared by many objectivists and subjectivists is that idealization will generate a unique result—a single ideal self—and that underdetermination is therefore not a pressing problem for the idealizing subjectivist. If the argument of this paper is correct, that assumption is false.

But I am not suggesting this is a reason not to idealize, rather, I am suggesting that further machinery is required for those troubled by the recurrence of the underdetermination problem at the level of ideal *selves*. But once we are operating at the level of equally ideal selves, I see no reason why the preferences of an actual agent *surveying those ideal selves* cannot be relevant to deciding who to be and which life to lead. By hypothesis, every option is equally rationally permissible—the field has already been radically trimmed by the requirement that each option be integrated into an ideally coherent evaluative perspective—so it seems right to me that, at this point, it is just “up to us”. Our options have been tidied-up in light of full information and the requirements of rationality.

So understood, the trusteeship solution to the underdetermination problem offers a degree of rational freedom that is not overwhelming, but empowering. This, in turn, gets the balance right between rationality and alienation. An animating concern of

<sup>16</sup> Cf. (Chang, 2009, 2013), who argues that we create voluntarist reasons where our given reasons run out. My view is similar to Chang’s in some respects, though, of course, I am not arguing that there are given reasons explained by facts about objective values, nor am I arguing that we create voluntarist reasons by choosing to take *individual* considerations as reasons, nor am I arguing that there are deontic constraints on what an agent can recognize as a reason. Cf. (Gert, 2018)’s criticisms of Chang.

<sup>17</sup> Thanks to an anonymous referee for pressing this objection.

subjectivism is to ensure that we are not alienated from our own good, while also not allowing our decisions to be dominated by false beliefs, incomplete information or irrationality. But a persistent concern with idealizing accounts of subjectivism is that our ideal self (or, if I am right, *selves*) would have to be so unlike us in order to weigh the various options that they are no longer really *us*. What they command “from on high” now looks too remote from our present concerns to be consistent with a key motivation for being a subjectivist in the first place. But the trusteeship solution answers this concern by placing the authority to decide what is best squarely on the shoulders of the non-ideal agent. We decide whom to *entrust* with our authority thereby preserving our freedom. We ensure that the authority is wise, because ideal and we ensure that it is *trustworthy* because, being us, it has our best interests at heart. Now, of course, there remains the possibility that the actual deliberating agent will find *all* the options prescribed by their ideal selves alienating. In this scenario, it may be of little comfort that the decision ultimately rests on their shoulders. But even in this scenario, it still seems a point in favour of the trusteeship solution that some choice between alternatives is left to the agent. Of course, it is better to have a choice between alternatives you find agreeable, but if the only alternatives available are disagreeable, it still seems better to have a choice among them than none at all, precisely because, if nothing else, the choice gives you ownership over your life in a way that respects your autonomy as a rational agent. Besides, one might add, in most cases selecting between ideal options it is not unreasonable to hope that your participation in selecting it, together with the knowledge that the option is genuinely ideal, might be enough to bring you around to it.

It is, of course, at least possible that there will be some agents for whom *every single one* of their ideal selves appears utterly remote *and* unappealing, and therefore alienating. For these agents, it may be small consolation that they get to choose between many remote and unappealing options. But such agents are likely to be very rare. Each of us will have many, many ideal selves and, since each ideal self is a function of the desires, preferences and values of the actual self, it would be surprising (to the say least) if *none* of them were at all *appealing* to us. Granted, they might seem *remote*, given how far short of ideal rationality we actual agents generally are. But for most agents, I suspect, this will *add* to the appeal of at least some of them, as they recognize their ideal counterparts for what they are: a better version of themselves. This, in combination with the normative and psychological significance of *choosing between them*, greatly reduces the alienating aspect of their residual remoteness.

Now, of course, a further objection might be that the trusteeship solution therefore defaults to conservatism. One might worry that *inevitably* the actual self will be biased to their “most proximate” ideal self. But I do not think this is true. Of course, some people will choose the self that is most familiar and does least violence to their existing plans and self-conception. Perhaps Greenie will be most tempted by the life of a climate scientist in New York after all. But this need not be the case. One can imagine Greenie surveying her candidate selves, contemplating the eco-terrorist and thinking, with a rush, ‘Wow, I never knew I had it in me. This is what I have been looking for all along.’

This point also helps us to see the significance of the following fact: it is precisely because the actual self would respond in a determinate way in the hypothetical choice scenario that we can resolve the subjectivist’s normative underdetermination problem

generally, i.e. determine the actual agent's normative reasons. If indeterminacy permeated the hypothetical choice as well, then the underdetermination problem would remain unresolved. According to the trusteeship solution, in the hypothetical choice scenario, the actual self appraises the various ideal selves and decides which to entrust with rational authority. This is not the same as surveying the possible actions prescribed by the ideal selves and choosing on the basis of the intrinsic appeal of each action alone, judged by the actual self's own lights. This is not to say that the actual self *could not* decide whom to entrust on this basis. If the desire to perform some action were sufficiently strong, that might rule out every ideal self as a candidate, bar one, namely, the one prescribing that action. But the point is that further information is available to the actual self in the decision scenario in question, namely, how that action fits into an overall plan for their life endorsed by a particular ideal self. So, that an actual agent would choose some action under non-ideal conditions does not entail they would choose it under conditions contemplating their ideal selves, even if it were available. This speaks to the normative significance of their choice under these counterfactual conditions, which is part of the rationale for idealizing in the first place. The ideal self mediates and organizes the considerations which may be marshalled in favour of pursuing some course of action by integrating them into a holistic evaluative standpoint.

Is the trusteeship solution to the subjectivist's underdetermination problem hopelessly abstract? I hope not. In fact, I believe the account vindicates a tidied-up version of how some reflective people already do deliberate in their everyday lives. This, in turn, suggests that a decision-procedure closely connected to the account is both possible (because actual) and attractive. Of course, actual people do not go around trying to get imaginatively acquainted with their ideal selves and choosing between them. Subjectivists never intended idealization to supply a decision-procedure: that would be to over-intellectualize the relationship between actual and ideal selves. But nor is something like this procedure completely alien to our actual normative practices. Anytime someone selects a role model; or turns down a high-powered job because they realise they do not want to become the person they would have to be to succeed in it; or picks up a self-help book encouraging them to 'be the best version of themselves'; or advises a friend against a course of action because 'you don't want to go down that road, you might not like who you'll become', they implicitly engage in the sort of reasoning I have suggested is key to solving the underdetermination problem. This is literally *self-conscious* deliberation, where one considers the impact of some course of action on who one will become as well as on its other merits. Naturally, much more would need to be said to actually develop a decision procedure that best approximated the trusteeship solution while correcting for our various biases and other cognitive limitations, but I believe that our present normative practices indicate that the prospect of such a procedure is both possible and appealing. Careful deliberation about what to do often requires attempting to integrate our reasons about what to do *now* into a coherent picture of how we want to live our life *generally* and over the long run. A very important heuristic we already employ to do this is to ask, 'What sort of person do I want to be?' and one way of asking and answering that question is by imagining different ways our lives might go and the different people we would need to be to make the best of them.

In the first half of the paper, I argued that underdetermination is just as big a problem for subjectivists as it is for objectivists. In the second half of the paper, I presented two solutions to the subjectivist's underdetermination problem. The first was democratic, which confronted both technical and principled objections. I argued that the technical objections could be solved, but that the principled objections remained; and, if anything, were exacerbated by the solutions to the technical objections. I then presented an alternative solution I labelled 'the trusteeship solution'. I argued this solution avoided the in-principle objections to the democratic solution and was independently attractive, given the subjectivist's other commitments. I concluded by observing that the trusteeship solution vindicates a tidied-up version of how some reflective people already do deliberate in their everyday lives. This, in turn, suggests that a decision-procedure closely connected to the solution is both possible (because actual) and attractive.

**Funding** The Gates Cambridge Trust.

**Availability of data and material** Not applicable.

**Code availability** Not applicable.

## Declarations

**Conflict of interest** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Arrow, K. (1963). *Social choice and individual values* (2nd ed.). Yale University Press.
- Black, D. (1948). On the rationale of group decision-making. *Journal of Political Economy*, 56, 23–34.
- de Borda, J.-C. (1784). Mémoire sur les élections au scrutin par M. de Borda in *Mémoires de l'Académie Royale des Sciences année 1781* (pp. 657–665). l'Imprimerie Royale.
- Brandt, R. (1979). *A theory of the good and the right*. Oxford University Press.
- Bratman, M. (1987). *Intentions, plans and practical reason*. Harvard University Press.
- Broome, J. (2001). Are intentions reasons? And how should we cope with incommensurable values? In C. W. Morris & A. Ripstein (Eds.), *Practical rationality and preference: Essays for David Gauthier* (pp. 98–120). Cambridge University Press.
- Callard, A. (2018). *Aspiration: The agency of becoming*. Oxford University Press.
- Chang, R. (2002). The possibility of parity. *Ethics*, 112, 659–688.
- Chang, R. (2009). Voluntarist reasons and the sources of normativity. In D. Sobel & S. Wall (Eds.), *Practical reason and action* (pp. 243–271). Cambridge University Press.
- Chang, R. (2013). Grounding practical normativity: Going hybrid. *Philosophical Studies*, 164, 163–187.
- Chang, R. (2021). What is it to be a rational agent? In R. Chang & K. Sylvan (Eds.), *The Routledge handbook of practical reason* (pp. 95–109). Routledge.

- Darwall, S. (1983). *Impartial Reason*. Cornell University Press.
- de Condorcet, M.J.A.N. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: l'Imprimerie Royale.
- Dorsey, D. (2017). Idealization and the heart of subjectivism. *Noûs*, 51(1), 196–217.
- Dovi, S. (2018). Political Representation. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/fall2018/entries/political-representation/>.
- Enoch, D. (2005). Why Idealize? *Ethics*, 115, 759–787.
- Gert, J. (2018). Underdetermination by reasons. In D. Star (Ed.), *The Oxford handbook of reasons and normativity* (pp. 443–460). Oxford University Press.
- Gibbard, A. (1990). *Wise choices, apt feelings*. Harvard University Press.
- Heathwood, C. (2005). The problem of defective desires. *Australasian Journal of Philosophy*, 83, 487–504.
- Holton, R. (2006). The act of choice. *Philosophers' Imprint*, 6, 1–15.
- Hubin, D. (1996). Hypothetical Motivation. *Noûs*, 30, 31–54.
- Kolodny, N. (2011). Aims as Reasons. In S. Freeman, R. Kumar, & R. J. Wallace (Eds.), *Reasons and recognition: Essays on the philosophy of T. M. Scanlon* (pp. 43–78). Oxford: Oxford University Press.
- Korsgaard, C. (2009). *Self constitution: Agency, identity, and integrity*. Oxford University Press.
- Lewis, D. (1989). Dispositional theories of value. *Supplementary Volume, Proceedings of the Aristotelian Society*, 63, 89–174.
- List, C., & Puppe, C. (2009). Judgment aggregation: A survey. In P. Anand, C. Puppe, & P. Pattanaik (Eds.), *The handbook of rational and social choice* (pp. 457–482). Oxford University Press.
- Mill, J. (2009). *Utilitarianism*. The Floating Press.
- Murphy, M. (1999). The simple desire-fulfillment theory. *Noûs*, 33, 247–272.
- Portmore, D. (2011). *Commonsense consequentialism: Wherein morality meets rationality*. Oxford University Press.
- Portmore, D. (2013). Perform your best option. *Journal of Philosophy*, 110, 436–459.
- Quine, W. (1976). Posits and reality. In W. Quine (Ed.), *The ways of paradox and other essays* (pp. 246–254). Harvard University Press.
- Quine, W. (1980). Two dogmas of empiricism. In W. Quine (Ed.), *From a logical point of view* (pp. 20–46). Harvard University Press.
- Railton, P. (1986a). Facts and values. *Philosophical Topics*, 14, 5–31.
- Railton, P. (1986b). Moral realism. *The Philosophical Review*, 95, 163–207.
- Raz, J. (1999). *Engaging reason*. Oxford University Press.
- Raz, J. (2012). *From normativity to responsibility*. Oxford University Press.
- Rosati, C. (1995). Persons, perspectives, and full information accounts of the good. *Ethics*, 105, 296–325.
- Sampson, E. (2021). What if ideal advice conflicts? A dilemma for idealizing accounts of normative practical reasons. *Philosophical Studies*, 179(4), 1091–1111.
- Sen, A. (2004). Incompleteness and reasoned choice. *Synthese*, 140, 43–59.
- Sklar, L. (1975). Methodological conservatism. *The Philosophical Review*, 84, 384–400.
- Smith, M. (1994). *The moral problem*. Wiley-Blackwell.
- Sobel, D. (2016). *From valuing to value: Towards a defence of subjectivism*. Oxford University Press.
- Southwood, N. (2019). Contractualism for us as we are. *Philosophy and Phenomenological Research*, 99, 529–547.
- Street, S. (2009). In defense of future Tuesday indifference: Ideally coherent eccentrics and the contingency of what matters. *Philosophical Issues*, 19, 273–298.
- Ullmann-Margalit, E., & Morgenbesser, S. (1977). Picking and choosing. *Social Research*, 44, 757–785.
- Walden, K. (2015). The Euthyphro Dilemma. *Philosophy and Phenomenological Research*, 90, 612–639.
- Williams, B. (1981). Internal and external reasons. In *Moral Luck*. Cambridge University Press.