Philos Stud
https://doi.org/10.1007/s11098-021-01688-z

What if ideal advice conflicts? A dilemma for idealizing accounts of normative practical reasons

Eric Sampson[1]

Accepted: 19 June 2021

Abstract One of the deepest and longest-lasting debates in ethics concerns a version of the Euthyphro question: are choiceworthy things choiceworthy because agents have certain attitudes toward them or are they choiceworthy independent of any agents' attitudes? Reasons internalists, such as Bernard Williams, Michael Smith, Mark Schroeder, Sharon Street, Kate Manne, Julia Markovits, and David Sobel answer in the first way. They think that all of an agent's normative reasons for action are grounded in facts about that agent's pro-attitudes (e.g., her desires, valuing states, normative judgments). According to the most popular brand of internalism, idealizing internalism, an agent's reasons are grounded, not in her *actual* pro-attitudes, but rather in what her pro-attitudes *would be* in suitably idealized conditions. Idealizing internalists presuppose that, for any agent with an irrational set of attitudes, there is one uniquely rational set that that agent would have if she were to undergo the relevant idealizing process. I argue that this assumption is false and that it raises two puzzles for idealizing internalism: one about the *existence* of practical reasons and another about their normative *weight*. I argue that idealizing internalists have an adequate solution to the first puzzle but not the second. Indeed, when they try to solve the second puzzle, they confront a dilemma. This second puzzle and the associated dilemma thus constitutes a powerful, but so far unnoticed, difficulty for idealizing internalism.

Keywords Reasons · Internalism · Normativity · Rationality · Coherence

✉ Eric Sampson
eric.sampson@georgetown.edu

[1] Georgetown University, Washington, D.C., USA

Published online: 04 July 2021

1 Introduction

One of the deepest and longest-lasting debates in ethics concerns a version of the Euthyphro question: are choiceworthy things choiceworthy because agents have certain attitudes toward them or are they choiceworthy independent of any agents' attitudes?[1] Reasons internalists, such as Williams (1981), Smith (1994), Schroeder (2007), Street (2008), Manne (2014, 2016), Markovits (2014), and David Sobel (2017) answer in the first way. They think that *all* of an agent's normative reasons for action are grounded in facts about that agent's pro-attitudes (e.g., her desires, valuing states, non-representational normative judgments).[2] Reasons externalists, by contrast, think that *none* of an agent's reasons[3] are grounded this way. Externalists think that all of an agent's reasons are either themselves normatively fundamental, and therefore not grounded in some deeper reality, or that they're grounded in non-psychological facts about, for example, stance-independent value, fittingness, or oughts.[4] In any case, externalists deny the key internalist thesis that all reasons are grounded in pro-attitudes.[5]

Internalism has been enormously popular in ethics and for good reason. Among its many attractions are that it promises to: fit comfortably with metaphysical naturalism, make sense of differing reasons ascriptions for different agents (e.g., it explains why Ronnie, who likes dancing, has a reason to dance, but not Bradley who hates it), make normative epistemology fairly straightforward, explain how reasons can explain behavior, avoid alienating agents from their reasons, and explain the close connection between reasons and good reason*ing*. Externalism, it's often said, can do none of these things.

While internalists generally agree that internalism enjoys (many of) these advantages, they disagree about how to formulate the view. They disagree specifically about whether the pro-attitudes grounding an agent's reasons are that agent's *actual* pro-attitudes or rather what that agent's pro-attitudes *would be* if she

[1] A pithier, but less precise, way to put it: are the truths about practical normativity invented or discovered?

[2] A quick note about terminology. Some (e.g., Schroeder 2007) reserve the label "reasons internalism" for the view that reasons must be, in some sense, motivating. They reserve the label "The Humean Theory of Reasons" for the view that an agent's reasons are grounded in facts about that agent's desires. If you're wondering how these views are any different, see Markovits (2014), where she argues that reasons are grounded in desires but aren't necessarily motivating (though she calls herself an internalist). The terminology in this literature is a bit of a mess, though everyone seems to understand what everyone means. I actually prefer Sobel's (2017) term "subjectivism" to capture the range of views I mean to target here. That's because all my targets say that an agent's reasons are grounded in facts about the motivational psychology of subjects. But most of my targets self-identify as internalists, and the others, such as Street, don't deny the label so much as use different ones that they prefer. So I've settled on calling them all internalists.

[3] By "reasons for action" or "practical reasons", I will always have in mind *normative* or *justifying* reasons for action (as opposed to *motivating* reasons for action).

[4] Defenders of externalism include Parfit (2011), Scanlon (1998), Darwall (1983), Enoch (2011), Shafer-Landau (2003, 2012), and Cuneo (2007).

[5] I'll mostly use "desires" throughout since that's largely how my interlocutors talk, but I mean to target any view according to which an agent's reasons are grounded in their pro-attitudes.

were, in some sense, better situated (e.g., better informed, more reflective, less akratic). Actual internalism—the view that an agent's reasons are grounded in her actual pro-attitudes—is the simplest version of the view, but it's commonly thought to give the wrong verdicts about cases.[6] Most internalists are convinced that agents can desire or value things that, in the end, aren't desirable or valuable at all. It seems clear, for example, that agents' desires can be ill-informed, artificially aroused (e.g., the result of slick advertising), morally base, or in some other sense defective.[7] Many internalists wish to reserve the right to say that those kinds of desires don't ground, or give rise to, reasons for action. Thus, according to the most popular brand of internalism, *idealizing internalism*, an agent's reasons are grounded, not in her actual pro-attitudes, but rather in what her pro-attitudes would be in suitably idealized conditions.

A typical idealizing internalist view says that an agent, A, has a reason to Φ iff (and because) she has an ideal counterpart, $A+$, that has some desire that would be served by A's Φ-ing. Many internalists call $A+$, the actual agent's ideal counterpart, an "ideal advisor", since $A+$ shares many of A's concerns and sensibilities, and (almost certainly) has desires about how A acts that keeps A's interests at heart. One popular way of constructing $A+$'s desires (which I take to be representative of others) is as follows. You begin with A's actual motivational set (or set of desires), endow A with full (relevant) non-normative information, rid A of any false beliefs, and make A's beliefs and desires perfectly structurally (or procedurally) rational. Attitudes are structurally rational when they cohere or "hang together" in all the appropriate ways. Spelling out what it is for attitudes to hang together appropriately is no easy task. But however it's spelled out, it will almost certainly rule out clearly incoherent states of mind such as holding inconsistent beliefs, believing contradictions,[8] intending to perform two actions the agent believes to be incompatible, and being akratic (i.e., judging that you ought to Φ all-things-considered, while having no intention to Φ). Note that internalists don't imagine that an ideal advisor is perfectly *substantively* rational—the kind of creature who perfectly responds to her reasons. That's because internalists, in appealing to the notion of an ideal advisor, are attempting to give an account of the source, or metaphysical grounding, of reasons. It would therefore be theoretically useless to appeal to the notion of a creature who perfectly responds to their reasons in the course of giving a metaphysical account of the source of reasons.

In what follows, I'll argue that, for many actual agents' motivational sets, an idealizing process appealing to the notion of structural rationality will not determine a uniquely rational set. In other words, many agents have multiple ideal advisors. This is because the standards of structural rationality are fairly permissive. There are many ways for one's attitudes to be coherent—to hang together in the right way. There are

[6] Schroeder (2007) is probably the most prominent defender of actual internalism.

[7] For a nice discussion of defective desires in the context of the well-being literature, see Heathwood (2008).

[8] Having inconsistent beliefs is not the same as believing a contradiction. An agent has inconsistent beliefs if she has two beliefs, one of which is *that* p and the other is *that* $\sim p$. An agent believes a contradiction if she has one belief *that* p *and* $\sim p$. Both states of mind are incoherent.

therefore many ways for an agent's irrational motivational set to be corrected so as to satisfy the demands of structural rationality, even after the full-information condition has been met. Call the claim that, for at least some agents with structurally irrational psychologies, there is no uniquely rational way of correcting them *Non-Uniqueness*.

If Non-Uniqueness is correct, it gives rise to two puzzles for idealizing internalists. First, when Non-Uniqueness holds for an agent, which of her many ideal counterparts grounds the facts about what reasons *there are* for her? I'll consider several ways for idealizing internalists to answer this question and argue that, while none is particularly attractive, there is one response that is clearly superior to the others. The second puzzle, however, is more difficult: When Non-Uniqueness holds for an agent, which of her many ideal counterparts grounds the facts about the normative *weights* of her reasons?[9] After all, we don't just want our theory of reasons to return verdicts about what reasons there are. We also want it to tell us what an agent has *most* (or at least sufficient) reason to do. But, to do this, we need our theory of reasons to return verdicts about the weights of reasons—which considerations count more or less toward justifying a course of action. This is where the real difficulty lies for idealizing internalism. I consider several proposals for solving the second puzzle on idealizers' behalf and argue, piecemeal, that each fails. But then I argue that idealizers face a deeper problem still: when they attempt to solve this second puzzle, they confront a dilemma. The dilemma, very briefly, is this: either idealizing internalists will have to pick (entirely arbitrarily) one of an agent's ideal counterparts and insist that the weights of her reasons are determined by *that* ideal counterpart (and not any of the other equally ideal counterparts), or they will have to insist that the strengths of the actual agent's reasons are grounded in a set of attitudes it would be *irrational* for the actual agent to have. But it just isn't plausible that the strengths of an agent's reasons are grounded in a set of pro-attitudes that it would be *irrational* for her to have. So, either horn is deeply unattractive for idealizing internalists. This second puzzle and the associated dilemma therefore constitutes a powerful, but so far unnoticed, challenge to idealizing internalism.

2 The case for non-uniqueness

Whether an idealizing process will result in a uniquely rational set of desires for some actual agent with an irrational psychology, rather than several equally rational desire sets, depends very much on what the proposed idealizing process is like. Williams (1981), Smith (1994), Manne (2014), Markovits (2014), and Sobel (2017) all defend different idealizing conditions.[10] However, Smith's are the most

[9] We might also speak of the *strength* of an agent's reasons. I will use the weight and strength metaphors interchangeably.

[10] Street doesn't appeal to ideal counterparts, but rather to what "follows from" an actual agent's normative judgments. This difference is not significant for my purposes. What matters is that, even for Street, it's not the agent's actual attitudes that ground her reasons, but rather some "cleaned up" version of her attitudes. In this sense, Street is an idealizer. And Street's view is like the others in that it doesn't guarantee that, for any irrational agent, there's a uniquely correct way to clean up her attitudes.

restrictive.[11] In other words, if any of the accounts of ideal conditions guarantees that, for any possible agent's irrational motivational set, there is one uniquely rational way to correct it, it will be Smith's account. I will therefore deal mostly with his conception of ideal conditions, and argue that, if Non-Uniqueness holds in his case, it will hold in the others as well. But I want to emphasize that I mean to target all of these views, not just Smith's. While the details will be a bit different in each case, my strategy is the same: take the idealizing conditions offered by the author, show that they do not guarantee a unique ideal advisor (since coherent attitudes can be achieved in many ways), criticize various proposals for fixing the problem piecemeal (since the same kinds of issues arise for each account), and then (just to pile on) pose a dilemma, which I develop toward the end of the paper.

According to Smith, an actual agent *A*'s reasons for action are metaphysically determined by facts about what her counterpart in ideal circumstances would desire that *A* do. Building on work from Williams (1981), Smith understands the ideal circumstances as follows:

1. The agent must have no false beliefs
2. The agent must have all relevant true beliefs, and
3. The agent must deliberate correctly (Smith, 1994: 156).[12]

[11] Williams, Street, Manne, and Sobel are all Humean internalists. They think that the standards of ideal rationality (in combination with full information) are not so restrictive so as to guarantee that any possible agent will have a reason to comply with the demands of morality. Smith and Markovits, by contrast, are Kantian internalists. They think that the standards of rationality (in combination with full information) are restrictive enough to guarantee that agents always have a reason to comply with the demands of morality. Thus, on their view, any agent who fails to comply with the demands of morality is guilty of some kind of incoherence or structural irrationality. Smith's idealizing conditions are more restrictive than Markovits's because Smith thinks that, in addition to being structurally rational, ideal agents must be rid of any depression, addiction, or compulsions, and that they must sufficiently imagine what it would be like to have their desires satisfied, whereas Markovits's conditions only require perfect structural rationality. Moreover, Smith thinks that ideally rational agents' desires must exhibit maximal unity. (More on this in the coming paragraphs.) These additional requirements from Smith may lead one to wonder whether he is indeed a Kantian internalist. After all, many think that what's distinctive about the Kantian project is that it's an attempt to derive moral obligations from full information and the standards of structural rationality alone. But it seems that Smith might be adding substantive constraints on rationality—constraints that go beyond the mere coherence requirements of structural rationality. There are others who can more ably settle this question of taxonomy. I'll simply note here that Smith's project is close enough to the Kantian project that I think it's worth discussing his view along with other Kantians. And I think that my criticisms of Smith's view are even more pressing for those philosophers who fit more comfortably in the Kantian camp.

[12] It's worth noting that Smith doesn't think that conditions (1) and (2) are demands of rationality, only demands of ideal agency. Condition (3) is the one that incorporates demands of rationality. On Smith's view, one deliberates correctly only if one perfectly complies with the demands of structural rationality (and perhaps more substantive rational requirements, too). As I mentioned in footnote 11, some of Smith's conditions for ideal rationality (e.g., the absence of depression) seem to move beyond mere structural rationality or coherence. (After all, what's incoherent about being depressed?) Thus, there is a danger that Smith cannot justify his conditions on perfect rationality without an implicit appeal to reasons. This would mean that his account of reasons is viciously circular, since it analyzes reasons in terms of reasons. I will assume here that Smith is not vulnerable to this charge, but I have my doubts.

Condition (1) is in place to rule out the possibility that an agent can have a reason to do something when she desires to do that thing only because she has a false belief. Suppose, for instance, that I desire to drink the contents of a glass because I believe them to be gin and tonic. As a matter of fact, however, the glass contains petrol.[13] Many (including Williams and Smith) think that it's a strike against a theory of reasons if it returns the verdict that I have a reason to drink the contents of the glass. Condition (1) secures this result, since my ideal advisor, who knows what is really in the glass, would not want me to drink its contents. Idealizing internalism therefore says that I have no reason to do so.

Condition (2) rules *in* the possibility that an agent can have a reason to perform some action if her lack of a desire to do so is the result of ignorance. Suppose, for example, that I desire to buy a Picasso painting.[14] I don't know it, but the pawn shop down the street has just acquired a genuine Picasso. But the owners of the shop don't know that it's a genuine Picasso either. They assume it's just some old painting. If I were to go down to the shop, I could easily buy the painting for the $20 price. Is there a reason for me to walk down the street to the shop? Intuitively, there is, since doing so promotes my desire to buy a Picasso. Including condition (2) in the idealizing conditions secures this result. My ideal advisor, who has all relevantly true beliefs, would desire that I walk down the street to the pawn shop. Idealizing internalism therefore says that I do have a reason to do so.

Condition (3) says that the agent must deliberate correctly. Again, following Williams (1981), Smith argues that to deliberate correctly an agent must have sufficiently imagined what it would be like to satisfy her desires. If, for example, an agent desires that she buy a lottery ticket because she wants to win the lottery, but she has not sufficiently imagined what it would be like to be a lottery winner—for example, she hasn't imagined that some of her familial relationships will sour as family members scheme to get her money—then she has not deliberated correctly. If she were to imagine sufficiently what it would be like to win the lottery, her ideal self would not (let us suppose) desire that she win the lottery. So, her ideal self would not desire that she purchase a lottery ticket. Idealizing internalism says (correctly, it seems) that the actual agent would therefore have no reason to buy a lottery ticket.

In addition to having sufficiently imagined what it would be like to satisfy her desires, an agent's desires must also, on Smith's view, be "systematically justifiable" (p. 159). A set of desires is systematically justified if it exhibits "unity" to a maximal degree. Smith writes,

> Suppose we take a whole host of desires we have for specific and general things; desires which are not in fact derived from any desire that we have for something more general. We can ask ourselves whether we wouldn't get a more systematically justifiable set of desires by adding to this whole host of specific and general desires another general desire, or a more general desire still, a desire that, in turn, justifies and explains the more specific desires that

[13] This is a famous example of Williams's (1981).

[14] This example is similar to Smith's (1994: 157).

> we have. And the answer may be that we would. For in so far as the new set of desires…exhibits more in the way of, say, unity, we may properly think that the new imaginary set of desires is rationally preferable to the old. For we may properly regard the unity of a set of desires as a virtue; a virtue that in turn makes for the rationality of the set as a whole (1994: 159).

Smith doesn't say much more about what the unity of a set of desires consists in, but perhaps we have a sufficient grasp on what he means. For instance, imagine an agent who desires to eat healthy, get plenty of exercise, and save twenty percent of her income for retirement, but who lacks any desire to flourish in her old age. While this agent's desires do not seem blatantly incoherent, they would be more unified, it seems, if she were to adopt a more general desire to flourish in her old age. Such a desire would justify and explain the more specific desires having to do with diet, exercise, and retirement saving. Or imagine a person who desires to live well into her nineties but also heartily enjoys, and therefore desires, thrilling experiences like skydiving, harnessless rock climbing, and wrestling crocodiles. Again, there doesn't seem to be any blatant incoherence in this agent's attitudes, but it does seem that her attitudes would exhibit more unity if she were to drop either her desire for longevity or her desire for such exhilarating experiences. While I have my doubts about whether this kind of unity is a mark of ideal rationality, I'm happy to grant, for the sake of argument, that ideally rational agent's exhibit this kind of unity to a maximal degree.[15]

Since Smith is mainly concerned with rationally justifiable desires, he doesn't say much about other requirements of structural rationality. But I am supposing that the usual ones are implied in his account. For instance, I assume that ideally rational agents satisfy.

> **The Non-Contradiction Requirement:** You are rationally required not to believe that $\sim p$, if you believe that p.
> **The Enkratic Requirement:** You are rationally required to intend to Φ, if you believe that you yourself ought to Φ,
> **The Means-End Requirement:** You are rationally required to intend to Ψ, if you intend to Φ and believe that in order to Φ you must Ψ, and.
> **The Transitivity Requirement:** You are rationally required to prefer A to C, if you prefer A to B and prefer B to C.[16]

[15] It seems to me that the agents I've just described recognize as good, and therefore desire, two things that bear an unhappy relation to one another—namely, being difficult to achieve jointly. But it's not a mark of irrationality (in my view, at least) to recognize as good, and therefore desire, things that are difficult for limited beings like us to achieve jointly.

[16] I've represented these requirements as narrow-scope requirements. That is, each requirement says of agents satisfying the antecedent that they are rationally required to adopt the particular attitudes in the consequent. Of course, one may also read them as wide-scope requirements, but wide-scope requirements are more permissive than narrow-scope requirements since wide-scope requirements permit agents to satisfy them *either* by adopting the attitude in the consequent *or* by dropping one of the antecedent attitudes. But since I'm trying to show that even on the strictest conception of structural rationality, Non-Uniqueness holds, I will proceed as if the requirements of structural rationality are narrow-scope (with the exception of the Transitivity Requirement for reasons explained in the main text).

In other words, I assume that ideally rational agents don't believe contradictions, act in accord with their conscience, take the necessary means to their ends, and avoid having intransitive preferences. So, on my interpretation of Smith's view, an ideally rational agent has no false beliefs, all relevantly true beliefs, maximally unified desires, and satisfies the Non-Contradiction, Enkratic, Means-End, and Transitivity Requirements.

Now we can ask: Could there be an agent that fails to meet Smith's conditions on ideal agency but for whom there is more than one way to correct their attitudes so as to satisfy those constraints? I think there can be, and often are, such agents. Imagine, for instance, an agent with intransitive preferences. She prefers A to B, B to C, and C to A. For this agent, there are (at least) three ways to correct this irrational state. The ideally rational ordering for this agent might go {A > B > C}, or {B > C > A}, or {C > A > B}.[17] Each candidate ordering eliminates the intransitivity of the actual agent's preferences and each ordering seems equally good from the actual agent's own perspective.

It might be thought that, in this case, the Transitivity Requirement should be read as a narrow-scope requirement such that it requires the agent to adopt a particular attitude—namely, a preference for A over C. In other words, one might insist that the Transitivity Requirement be read as follows: You are rationally required to [prefer A to C], if you prefer A to B and prefer B to C, where the portion in brackets is the attitude that is rationally required of you. This is the narrow-scope reading of the requirement because "rationally required" takes narrow scope over the conditional. Narrow-scope interpretations of rational requirements are often contrasted with wide-scope interpretations. The wide-scope interpretation of the Transitivity Requirement reads: You are rationally required to [prefer A to C, if you prefer A to B and prefer B to C], where the portion in brackets is the attitude that is rationally required of you. Here, "rationally required" takes wide scope over the conditional. An agent complies with the wide-scope Transitivity Requirement either by preferring A to C, dropping her preference for A over B, or dropping her preference for B over C. Thus, wide-scope requirements tend to be more permissive than narrow-scope requirements since they allow several ways for an agent to fulfil the relevant requirement. Narrow scope-requirements, on the other hand, tend to be more restrictive because they usually permit only one way of satisfying the relevant requirement.[18] So, it might be thought that the example above—of an agent's having intransitive preferences—doesn't successfully show that there are multiple ways of complying with the Transitivity Requirement because we might read it as a narrow-scope requirement. This would mean that there is only one way to comply with the requirement—namely to adopt a preference for A over C.

[17] Here "X > Y" reads "X is preferred to Y".

[18] However, Lord (2014), a prominent defender of narrow-scope requirements, accepts that even with narrow-scope requirements agents are rationally permitted to satisfy them in many ways. On Lord's view, narrow-scopers and wide-scopers agree about all the deontic facts—that is, they agree about which states of mind are rationally (im)permissible. What distinguishes the views, according to Lord, is their *explanations* of those deontic facts.

This thought, however, is mistaken. Since the assignment of particular variables ("A", "B", or "C") to particular options is arbitrary, the Transitivity Requirement will never require an agent to adopt a particular attitude or a particular preference. For instance, suppose an agent prefers cookies to ice cream, ice cream to cake, and cake to cookies. Which dessert ought to be assigned the "A" variable, the "B" variable, and the "C" variable? Couldn't we perfectly legitimately assign each variable to each dessert? Of course we could. And there's nothing special about desserts that contributes to this result. The same reasoning applies to any case of intransitive preferences. If so, then, for any instance of intransitive preferences, the Transitivity Requirement won't dictate a unique way to correct the intransitivity. The requirement will tell us only that the intransitivity must be corrected somehow. Thus, the Transitivity Requirement is necessarily wide-scope.[19]

To see how the possibility of intransitive preferences shows that irrational agents can have multiple ideal counterparts, consider the following case. Imagine a third-year college undergraduate, Laura, deciding which career to pursue: doctor, lawyer, or philosopher. She would like to be a doctor because she'd like to help people, make great money, and have an intrinsically interesting job, but she doesn't like the inflexibility of the work schedule or the large time investment that is graduate school. She'd like to be a lawyer because she likes the great pay, the relatively short graduate school time investment, and the challenge the work presents, but she doesn't find the job intrinsically interesting and she knows the work hours will be brutal. Finally, she'd like to be a philosopher because she finds it intrinsically interesting, likes the flexible work hours, and likes all the opportunities for travel, but she doesn't like the relatively low pay and the large time investment that is graduate school. When Laura considers whether to be a doctor or a lawyer, she prefers to be a doctor. When she considers whether to be a lawyer or a philosopher, she prefers to be a lawyer. But when she considers whether to be a philosopher or a doctor, she prefers to be a philosopher. Laura has intransitive preferences.

[19] What about the suggestion that, when an agent has intransitive preferences, she is under three conflicting narrow-scope requirements. That is, her preferences for $A > B$ and $B > C$ require her to prefer $A > C$, her preferences for $B > C$ and $C > A$ require her to prefer $B > A$, and her preferences for $C > A$ and $A > B$ require her to prefer $C > B$? If this is correct, then no matter what this agent does, she will be guilty of some rational failing. If, for example, she satisfies the requirement to prefer $A > C$, she'll violate the requirement to prefer $B > A$. If she satisfies the requirement to prefer $B > A$, she'll violate the requirement to prefer $C > B$. And so on. The question, then, is: what would this suggest about the agent's ideal counterparts? It seems that there are only two plausible options. Either the actual agent has no ideal counterparts (since there is no way of correcting this agent's irrational preferences such that she is not guilty of some irrationality), or she has several ideal counterparts (since complying with one of her three obligations is no more rational than complying with one of her other obligations). If she has no ideal counterparts, then, according to idealizing internalism, she has no reasons for action at all. But this is implausible. An agent's reasons for action don't disappear as soon as she has intransitive preferences. So it must be that agents with intransitive preferences have several ideal counterparts. If so, that would vindicate the claim being argued for in this section: that there are possible agents with multiple ideal counterparts.

Which is the ideally rational preference ordering for Laura? Is it {doctor > lawyer > philosopher}, {lawyer > philosopher > doctor}, or {philosopher > doctor > lawyer}?[20] The Transitivity Requirement, on its own, won't help us here. It just tells us that Laura's preferences must somehow be transitive. It doesn't tell us how Laura ought to go about achieving that result. It might be thought that the other idealizing conditions suggested by Smith will, by themselves, eliminate the intransitivity of Laura's preferences. For instance, it might be thought that once we rid Laura of all her false beliefs, endow her with full-relevant information, give her sufficient imaginative familiarity with the various outcomes, and try to fit Laura's desires into a maximally unified whole, that a uniquely rational preference ordering *must* emerge. But it's not clear why this should be so. The intransitivity of Laura's preferences need not depend on a false belief, or a lack of information, or a failure of imagination. Laura's scenario is perfectly intelligible if we suppose that she has all the relevant information and is perfectly imaginatively acquainted with all the options. The intransitivity of her preferences persists, we might imagine, because she regards the good-making features of each career as incommensurable (or, at any rate, she doesn't know how to go about comparing them). It's no doubt true that Laura's desires lack a kind of unity, so we can say that she fails to have maximally unified desires. But what we'd like to know now is what the uniquely unified set of desires for her is. What if there are three different maximally unified sets of desires for Laura?

If there are three different rationally permissible preference orderings for Laura, then there are (at least) three different rationally permissible sets of desires for her. Suppose that tomorrow is the deadline for Laura to sign up for the MCAT, the LSAT, or the GRE—three exams necessary for entrance into graduate school for being a doctor, a lawyer, and a philosopher respectively. She can only choose one. And suppose that she must also decide whether to ask advice about graduate school from one of three relatives: her aunt who is a pediatrician, her uncle who is a prosecutor, or her cousin who is a philosopher. Finally, suppose that she must decide which sites to browse in the coming months as she prepares to apply to the graduate schools of her choice: med school sites, law school sites, or philosophy program sites. If it is rationally permissible for Laura to correct her intransitive preferences by putting either of the three careers at the top of the preference ordering, then the result is that there will be three different sets of desires—corresponding to the three different sets of behaviors associated with the three professions—that are rationally permissible.

For instance, imagine that Laura corrects her intransitive preferences by putting being a doctor at the top of her preference ordering, followed by being a lawyer, and then a philosopher. If so, then, plausibly, the most rational set of desires for Laura is one containing the desires to sign up for the MCAT, speak to her pediatrician aunt, browse medical school sites, while desiring *not* to sign up for the other tests, *not* to

[20] Of course, there are more options here, like {philosopher > lawyer > doctor}, {lawyer > doctor > philosopher}, and so on. I'm simplifying the example to keep it manageable, since I will return to it throughout the paper. But it's worth keeping in mind that Laura's situation is actually more complicated than I let on in the main text, which makes Non-Uniqueness even *more* plausible.

seek advice from the other relatives, and *not* to browse the other graduate school sites (since doing so would frustrate, rather than serve, her aim to become a doctor).[21] If, however, Laura corrects her intransitive preferences by putting being a lawyer at the top of her preference ordering, followed by being a philosopher, and then a doctor, then a different set of desires will be most rational. It will be the set containing desires to sign up for the LSAT, seek advice from her prosecutor uncle, browse law school sites, while desiring *not* to sign up for the other tests, *not* to seek advice from the other relatives, and *not* to browse the other graduate school sites. The sets of desires that her ideal counterparts could have can be represented in the table below.

Most Preferred Career	Desire to take	Desire to seek advice from	Desire to browse
Doctor	MCAT, not others	Aunt, not others	Med School Sites, not others
Lawyer	LSAT, not others	Uncle, not others	Law School Sites, not others
Philosopher	GRE, not others	Cousin, not others	Phil Program Sites, not others

If this is correct—if, that is, there are at least three distinct sets of desires that are rationally permissible for Laura—then there is no uniquely rational set of desires for her. This would vindicate Non-Uniqueness—the claim that at least some irrational motivational sets are such that there is no uniquely rational way of correcting them.

3 A puzzle about the existence of reasons

If, as I've argued, Non-Uniqueness is correct, things get complicated for idealizing internalism. For the following question arises: When an agent has multiple ideal advisors, which one grounds the facts about what reasons there are for her?

Return to Laura's case. She has several ideal advisors, but there's no reason to think that one of them, and not the others, is especially well-placed to ground the facts about what reasons there are for Laura. After all, each counterpart is a fully-informed, perfectly-rational version of her. So it would be arbitrary simply to pick one and insist that *that* ideal counterpart, and not the others, determines the facts about Laura's reasons for action. There's nothing special about any of the counterparts that would justify this.

We might suggest that *if* Laura were to undergo the idealizing process, then she *would have* adopted one of the ideally rational sets of desires—even if it would have been rationally permissible for her to adopt one of the other sets as well. The

[21] I am assuming that, once Laura has set her career preference, she is rationally required to adopt the desires to take the corresponding test, seek the corresponding advice, and browse the corresponding sites. Thus, I assume, as I said in footnote 16, that the standards of rationality are narrow-scope, requiring Laura to adopt particular attitudes. Again, I'm comfortable with this assumption because narrow-scope requirements are less permissive. If I interpret the constraints of rationality as wide-scope requirements, then the task of vindicating Non-Uniqueness is easy.

thought here is that there are only three rationally permissible sets of desires for Laura. Each set is distinct, so it's logically impossible for Laura to simultaneously adopt all three desiderative profiles. For instance, if Laura's ideal counterpart prefers most for Laura to be a doctor, then she'll desire that Laura take the MCAT. But the ideal counterpart that prefers for Laura to be a lawyer has no such desire (and indeed desires that Laura *not* take the MCAT). So no agent could adopt the desiderative profile of all Laura's ideal counterparts. Nevertheless, if Laura does in fact undergo the idealizing process, she's bound to satisfy one and only one of those desiderative profiles. We might suggest, then, that whichever desiderative profile Laura would have adopted had she undergone the idealizing process is the one that grounds Laura's reasons for action now.

But this suggestion is only slightly better than picking one of the ideal counterparts and insisting that *that* counterpart, and not the others, is the one that determines what reasons for action there are for Laura. Suppose it's true of Laura that, if she were to undergo the idealizing process, she would adopt the desiderative profile of her ideal counterpart that most prefers for Laura to be a doctor, but that it would have been rationally permissible for her to adopt one of the other desiderative profiles of her ideal counterparts. Now suppose that actual Laura proceeds to act in all the ways advised, or licensed, or desired by the counterpart preferring that Laura be a philosopher. Laura signs up for the GRE, seeks advice from her philosopher cousin, and begins browsing the websites of various philosophy programs. What could be the complaint against Laura's behavior? She is, after all, acting in ways that are licensed, or advised, or desired by an ideally-informed, ideally-rational version of herself. In other words, she's acting in a way that is fully endorsed by someone who perfectly embodies her very own evaluative perspective. And it's this property—that one's ideal counterpart perfectly embodies one's evaluative perspective—that internalists have always thought makes ideal counterparts normatively authoritative. So, intuitively, it just isn't plausible that Laura is making any normative mistake when she acts contrary to the advice of the ideal advisor who wants her to be a doctor (but perfectly in line with the advice of the counterpart that wants her to be a philosopher). But this is precisely what the proposal under consideration says. It says that Laura has uniquely most reason to be a doctor and is therefore making a mistake by becoming a philosopher. And that's false.

One might attempt to defend internalists from the objection in the previous paragraph with the following thought. Internalists are, right from the start, committed to the view that our particular desires make a difference to our reasons—even though we could have permissibly had different desires, and thus different reasons. So, the thought goes, the view that the facts about which career Laura *would* prefer if she had undergone the idealizing process determines the facts about Laura's reasons (though the other options would have been rationally permissible too) is not so different from the view that the facts about what Laura desires determines the facts about Laura's reasons (though other desires would have been

permissible too).[22] If so, then the objection in the previous paragraph would lose much of its force.

There is, however, a relevant difference here. According to reasons internalists, when Laura contingently adopts a desire D1, and not some other antecedently permissible desire D2, D1 gives Laura reasons for action, not D2. That's because, once adopted, D1, and not D2, (partially) constitutes Laura's evaluative perspective. That is, once adopted, D1 embodies what Laura values or cares about. However, when it's true of Laura that, if she were to undergo the idealizing process, she would contingently adopt the preference for the life of a doctor, rather than the other permissible alternatives, it's still true of Laura that the alternative counterparts perfectly embody her evaluative perspective—they perfectly embody what Laura cares about. In this way, the cases differ. Once D1 rather than D2 is adopted, D2 no longer embodies Laura's evaluative perspective. But once it's true of Laura that she would prefer the life of the doctor if she underwent the idealizing process, the alternative counterparts still do, even in that moment, perfectly embody Laura's evaluative perspective. So, by internalism's own theoretical commitments, those counterparts' desires are normatively authoritative for her.

No doubt the tricky theoretical problem generated by Non-Uniqueness appears to go away if one insists that, even though Laura's other counterparts perfectly embody her evaluative perspective, she is nevertheless acting contrary to her reasons when she acts in accord with the desires of one of her other counterparts. But it's an otherwise unmotivated response. The only thing this proposal has going for it is that it seems to get internalists out of an awkward theoretical jam. The proposal is therefore ad hoc. Moreover, it seems to me that any intuitive plausibility this proposal has is due to the inchoate thought that the counterpart whose desires Laura would have adopted had she undergone the idealizing process *really does* embody Laura's evaluative perspective better than the other counterparts. By stipulation, however, this is not so. Each counterpart equally embodies Laura's evaluative perspective. So it's not clear why they would not equally ground Laura's reasons for action.

In my view, the best response idealizing internalists have to this puzzle is to insist that an actual agent's reasons are determined by *all* of her ideal counterparts' desires. The view, then, would be that an agent A has a reason to Φ just in case *at least one* of A's ideal counterparts has a desire that would be served by A's Φ-ing. So, in Laura's case, this view would entail that Laura has *a* reason to perform all the actions that are desired by her ideal counterparts. Laura therefore has *a* reason to sign up for the MCAT, the LSAT, and the GRE. She has *a* reason to seek advice from her aunt the pediatrician, her uncle the prosecutor, and her cousin the philosopher. And she has *a* reason to browse med school, law school, and philosophy program sites. This is because at least one of Laura's ideal counterparts desires that Laura do each of these things. This adjustment to standard idealizing internalism makes sense of the thought, often insisted upon in the reasons literature, that reasons come cheap. It is often said that there are reasons for agents to do all

[22] I owe this suggestion to Alex Worsnip.

sorts of things—even bizarre things like eating your car.[23] Eating your car would, after all, help you get your daily dose of iron. If an agent's reasons are determined by all of her ideal counterparts, then an agent will likely have all kinds of reasons—even bizarre reasons. Moreover, this adjustment to standard idealizing internalism allows idealizing internalists to retain their insistence that what reasons there are for an actual agent are determined by the facts about what their ideal advisors would advise or desire them to do in their actual circumstances. It's just that there can be multiple ideal advisors offering advice.

4 A puzzle about the weight of reasons

According to the suggestion in the previous section, idealizing internalists should embrace the view that the facts about what reasons there are for an agent are grounded in the desires of all that agent's ideal counterparts. If this view is correct, then an agent can have many reasons to perform many different, even incompatible, actions. But we wouldn't say that an agent has *most* reason to do everything she has *a* reason to do, and we certainly wouldn't say that she has most reason to perform incompatible actions. That's because what an agent has most reason to do depends not only on the existence of her reasons but also on their normative weights—how much each reason counts in favor of the agent's taking each course of action.

4.1 Idealizing proportionalism

How, then, are the weights of an agent's reasons determined when she has multiple ideal counterparts? Idealizing internalism, on its own, doesn't answer this question since it's a theory about the existence of practical reasons, not their weights. But reasons internalism, and therefore idealizing internalism, is almost always paired with a very natural theory of normative weight: Proportionalism.[24] According to Proportionalism, the weight of a reason R for an action A is proportional to two things: (1) the strength of the agent's desire that explains the existence of R, and (2) the degree to which A promotes the agent's desire. To borrow Schroeder's (2007) example, Proportionalism says that the strength of Ronnie's reason to go to the party is proportional to the strength of Ronnie's desire to dance and the degree to which going to the party would promote Ronnie's desire to dance. All things being equal, the stronger Ronnie's desire to dance, the stronger his reason to go to the party; and all things being equal, the more likely it is that there will be dancing at the party, the stronger Ronnie's reason to go. Proportionalism is attractive because it offers a powerful and unified theory of reasons: the existence of a reason is explained by the

[23] This example comes from Schroeder (2007).

[24] Schroeder (2007: Ch. 7) coined the name and first made the view explicit, though he rejects it. To my knowledge, he is the only internalist who rejects proportionalism. The view has been defended explicitly by Evers (2014), Manne (2016), Rieder (2016), and Sobel (2017: Ch. 15), and is implicitly endorsed by Smith (1994: 144).

existence of a desire, and the strength of a reason is explained by the strength of a desire.

Just what the strength of a desire amounts to is something about which proportionalists can disagree. The strength of a desire might be understood as its *phenomenological intensity* or "nagging", or "pull".[25] Alternatively, one might adopt Manne's (2016) view according to which a desire's strength is the *depth* of the desire in the agent's psychic economy. The depth of a desire is understood as its priority over the other desires the agent has. According to Manne, the relevant question to ask, if you want to know which of two desires is deeper, or which has priority in an agent's psychic economy, is "When a genie now appears and grants [a subject] S *n* wishes, where *n* remains unknown to S, what would be the order in which S would cast her wishes, all things being equal?" (Manne, 2016: 134).[26] I won't try to settle this intramural debate about what the strength of a desire amounts to for proportionalists. My criticisms will apply no matter which view proportionalists take. Manne's view strikes me as the most plausible,[27] so I will assume that the strength of a desire is best understood as its depth in an agent's psychic economy.

Combining Manne's brand of proportionalism with idealizing internalism yields the following view about the weight of practical reasons:

> **Idealizing Proportionalism:** The weight of a reason R for an agent A to perform some action Φ is proportional to: (1) the strength of A's ideal counterpart, $A+$'s, desire that explains the existence of R, and (2) the degree to which A's Φ-ing promotes $A+$'s desire.

On this view, Ronnie's reason to go to the party is proportional to the strength of Ronnie+'s desire for Ronnie to dance and the degree to which Ronnie's going to the party would promote Ronnie+'s desire that Ronnie dance. All things being equal, the stronger Ronnie+'s desire that Ronnie dance, the stronger Ronnie's reason to go to the party; and all things being equal, the more likely it is that there will be dancing at the party, the stronger Ronnie's reason to go.

[25] One might like a more precise characterization of what a desire's phenomenological intensity, nagging, or pull amounts to. I don't have such an account, but that is because no proportionalist has offered one—not even Rieder (2016) in his recent defense of the view. Rieder's main concern is to defend proportionalism against Schroeder's (2007) criticisms. However, all of Rieder's examples about the strength of a desire (e.g., his desire to drink coffee, his desire to keep writing in the face of a deadline) are examples in which the strength of the desire amounts to the phenomenological intensity of the desire. They are examples in which the phenomenological character of the desire is present to the agent's mind.

[26] I take it that Manne intends this only as a rough *test* for priority, not what priority consists in.

[27] I'm more or less convinced by Manne's (2016) criticism of the phenomenological intensity conception. Manne says that the strength of an agent's desire—construed as its phenomenological pull—can diminish when the strength of the agent herself diminishes. But this doesn't entail that the strength of the agent's reasons diminish. Manne considers a person who is starving or freezing to death. Often, when a person is starving or freezing, her sense of being cold or hungry tends to wane. So the more her bodily need for food or shelter increases, the more her sense of that need—the phenomenological pull she experiences—decreases. And this isn't the result of any incoherence or irrationality on the agent's part. So the implausible prediction that, as a person gets closer to freezing to death, they have weaker and weaker reason to find warmth, cannot be corrected by adopting idealizing proportionalism.

We can now begin to see the puzzle about the weight of reasons for idealizing internalists: When Non-Uniqueness holds for some agent, which of her many ideal counterparts grounds the facts about the weights of her reasons?

4.2 Adding up strengths

Perhaps our answer to the first puzzle can guide us in answering this one. In response to the first puzzle about the existence of an agent's reasons, we said that an agent has a reason to Φ just in case at least one of her ideal counterparts has a desire that would be promoted by her Φ-ing. But it's not immediately clear how to extend this answer to the present puzzle. Each of Laura's ideal counterparts has a desire about, say, whether Laura takes the MCAT. One desires that she take it (since it's necessary for becoming a doctor), while two desire that she *not* take it (since Laura can only take one test, and taking the MCAT would frustrate, rather than serve, their desires for Laura to become a lawyer or a philosopher). So, according to one ideal counterpart, Laura has strong reason to take the MCAT and no reason not to. But, according to two ideal counterparts, she has strong reason not to take the MCAT and no reason to take it. Perhaps, then, we should say that Laura has strong reason to take the MCAT and strong reason not to take it, since she has at least one ideal counterpart that desires strongly that she take each course of action. And that seems like the intuitively correct result. But then we can ask whether, all things considered, Laura has stronger reason *in favor* of taking the MCAT or stronger reason *against* taking it. We might answer this question by adding up the strengths of Laura's reasons to take the MCAT and adding up her reasons not to, according to each counterpart. This would entail that Laura has stronger reason *against* taking the MCAT than *for* it since the strengths of her two counterparts' desires that she not take it would outweigh the strength of her one counterpart's desire that she take it. This might sound like a good way to integrate the strengths of different ideal counterpart's desires so that Idealizing Proportionalism can return plausible verdicts about the strengths of reasons.

But this solution is bound to fail. If we continue in the way described above by adding up the strengths of all three counterparts, the result, according to Idealizing Proportionalism, will be that, for each test, Laura has stronger reason *against* taking it than *for* taking it. For it is true of each test that two of Laura's ideal counterparts desire strongly that she *not* take the test (since doing so would frustrate this counterpart's desire that Laura pursue the career that they want her to pursue) while one desires strongly that she take it (since doing so is necessary for Laura to pursue the career they want for her). The same is true of Laura's reasons to seek advice from her relatives. Two ideal counterparts will advise Laura not to seek advice from her pediatrician aunt, her prosecutor uncle, and her cousin the philosopher, while only one will advise her to seek advice from each. According to a straightforward application of Idealizing Proportionalism, then, the result will be that Laura has all-things-considered stronger reason *against* seeking advice from each relative than *for* seeking advice from them. But these are clearly implausible verdicts about Laura's reasons. Surely, when all her reasons are weighed up, she has stronger reason for, rather than against, pursuing at least one of her career options and to take the

corresponding steps concerning tests, advice, and browsing websites. These are, after all, her most preferred career options. So any plausible version of reasons internalism should return the correct verdict about the valence of Laura's all-things-considered reasons to pursue at least one of these careers.

It's irrelevant that, for each ideal counterpart, they prefer that Laura take one of the tests rather than none of the tests. The ideal counterparts aren't getting together and negotiating about what Laura will do in light of the advice that the other counterparts have given. Each counterpart is advising Laura in isolation, offering her advice about what to do from their own perspective. So, for each test, two counterparts will tell her emphatically "Don't take that exam!" (since it's incompatible with their desires for her) while one will tell her "Take that exam!" Our puzzle here is how to integrate this conflicting advice so that it returns plausible verdicts about Laura's reasons.

4.3 Equally strong reason

We might be tempted to say that Laura has all-things-considered equally strong reason in favor of pursuing each career option and equally strong reason to take the corresponding steps concerning tests, advice, and website browsing. After all, each career option enjoys an equal amount of favor and disfavor by the ideal advisors. (Though it's still true that each is more disfavored than favored.) And if we step back from theorizing for just a second, the idea that Laura has equally good reason in favor of pursuing each career seems like the intuitively correct result. She has three great options. It should turn out that she has equally strong all-things-considered reason in favor of pursuing them. But of course, we need the theory to deliver that result (and in a non-ad-hoc way). We can't just stipulate that it does. The trouble is that there's no straightforward way for the theory to do this.[28]

Notice that, on the present suggestion, while each career gets the same amount of support by the ideal advisors, this won't deliver the result that Laura has all-things-considered equally strong reason *in favor* of each career. Again, it's the opposite. As we said, each career choice is equally *disfavored*. Since each career choice is opposed by two advisors and favored by one, Idealizing Proportionalism delivers the bizarre result that, when it's all said and done, Laura has stronger reason not to pursue each career than to pursue them. It's no doubt true that each career is equally disfavored. But that's not good enough to deliver the intuitively correct result. Reasons have valences. They can be *for* or *against* taking courses of action. And Idealizing Proportionalism can't get the valence of Laura's all-things-considered reasons right.

[28] It's worth noting that externalism, according to which the weights of reasons are determined by the amount of value to be realized by each course of action *can* deliver this result. Externalists can say that, since each career is equally valuable, Laura has equally good reason *in favor* of pursuing each career. Alternatively, they might say that pursuing one of the careers is best, since it realizes the most value, but that, given the options' similarity in value, it's really difficult for Laura to *know* which one is most valuable. In any case, the fact that Laura has intransitive preferences poses no problem at all. According to externalists, reasons, and therefore the weights of reasons, are entirely independent of Laura's attitudes.

One way of hearing the suggestion that Laura has equally strong reason in favor of pursuing each career is that she has a fourth ideal advisor that I failed to mention. This advisor is in favor of all three options, but doesn't oppose any of them, and doesn't favor one career more than another. Moreover, the suggestion goes, this advisor is the uniquely correct ideal advisor for grounding the strengths of Laura's reasons. This suggestion solves the problem of getting the wrong valence for Laura's all-things-considered reasons to pursue each career because this ideal advisor equally favors each career and disfavors none of them. The trouble, of course, is that it confronts the arbitrariness problem: Why is this equally-favoring, not-disfavoring, advisor the uniquely correct advisor for grounding the strengths of Laura's reasons when three other equally well-informed, equally-rational ideal advisors exist? And those other three advisors don't favor each career equally. It seems the only rationale for insisting on this fourth advisor is that that's the one that solves the theoretical problem the view confronts—again, it's ad hoc. And the same is true for any other proposed counterpart that can be cooked up to solve the problem I've identified here. Any counterpart you think I've neglected will still be just one of a host of equally ideal counterparts for Laura. So we won't be able to solve this problem by identifying yet another counterpart for Laura and insisting that *that* one solves the problem.

4.4 Averaging

Since we get the wrong result about the valence of Laura's all-things-considered reasons, it won't help to suggest that we *average* the strengths of Laura's ideal counterparts' desires to determine the strengths of her reasons. For, even if the averages turn out to be equal to one another, the average strength of her advisors' desires *against* taking the MCAT (LSAT, or GRE) will still be stronger than the average strength of the counterparts' desires *in favor* of taking the MCAT (LSAT, or GRE), since, for each test, two ideal advisors will advise Laura by saying (or desiring) "Don't take the test!". The result, according to Idealizing Proportionalism, then, is that Laura has all-things-considered stronger reason not to take each test than to take it (though she has some reason to take each).

4.5 A committee of advisors

Here's a suggestion that may have occurred to you at the end of Sect. 4.2. Perhaps each ideal advisor would know that they are but one of many advisors offering advice to the actual agent and would therefore form a committee to deliberate about, and agree upon, what advice to give the actual agent. On this suggestion, the advisors would see the predicament I've been describing, realize that unless they agree on a recommendation there will either be no fact of the matter about the actual agent's reasons or that the all-things-considered valence of her reasons will be

intuitively wrong,[29] and therefore agree to advise the agent to just pick a career. They'd tell her "Each of the three careers is equally good." If so, then the ideal advisors would be in perfect agreement. Moreover, they would deliver the intuitively correct result. So none of the problems I've described would arise.

Here I begin to get philosophical vertigo. For this suggestion to work, these advisors would have to be aware that their own verdicts are the metaphysical grounds of their actual self's reasons. They would, moreover, need to see that, unless they come to agreement, the problems I've described so far will arise, they'll need to think that this is a bad thing, and therefore be motivated to try to come to agreement with one another. Otherwise, they'd simply issue their original advice to the agent and be done with it. But let's assume for the moment that this is all philosophically above board.

Even so, for any recommendation the committee of advisors gives, we could always ask "Could the committee have rationally settled on some *other* advice?" And the answer always seems to be "yes." They could have drawn straws so that one advisor wins and gets to issue the advice. This seems fair. They could have arm wrestled, or rolled dice, or played rock, paper, scissors, for the chance to give advice. In short, they could have adopted *any* impartial procedure for coming to agreement. So long as they decide to issue the advice given by at least one of the advisors, the advice would have embodied the evaluative perspective of the actual agent as well as any of the alternatives. But this means that any recommendation made by a committee is *arbitrary* from a rational point of view. There's no reason the committee's recommendation, whatever it turns out to be, should be any more normatively authoritative for the agent than the advice issued by each of her ideal advisors individually. For the committee could have permissibly chosen to side with any of them. Thus, the arbitrariness problem returns.

4.6 A dilemma

We could continue offering, and criticizing, proposals for how to solve the problems I've been discussing. I don't claim to have considered all possible options here. But I hope I've given a sense of the kinds of problems any such proposal is bound to face.

There is, however, a deeper problem for idealizing internalism. Even if a proposed tweak to idealizing internalism avoids some of the problems I've described above, such a proposal will face the following dilemma. The proposal will have to say either that the strengths of an agent's reasons are grounded in the strengths of the desires of one, and only one, of her many ideal counterparts, or they aren't. If the strengths of an agent's reasons *are* grounded in the desires of one, and

[29] What's so bad about this? Well, there clearly is a fact of the matter about Laura's reasons. For instance, she has good reasons in favor of pursuing each career. That is very much determinate. She has more reason to pursue these careers than to jump off a bridge. The fact that being a lawyer would help her pay her rent is a stronger reason to be a lawyer than the fact that, as a lawyer, she would occasionally get to speak in a microphone. These are determinate facts about her reasons and the theory better deliver that result.

only one, of her many ideal counterparts, then idealizing internalists confront the now-familiar arbitrariness problem: Why are the strengths of the actual agent's reasons grounded in *that* ideal counterpart's desires rather than any of the other equally ideal counterparts' desires? All the other counterparts have the property that idealizing internalists think makes an ideal counterpart normatively authoritative: they're ideally-informed and ideally-rational. To insist on one counterpart, excluding all others, is arbitrary.

If, however, the proposal says that the strength of an agent's reasons are *not* grounded in the strengths of one, and only one, of her ideal counterpart's desires, but rather some amalgamation of them (e.g., the sum, the average), then it is difficult to see why the actual agent's reasons would be grounded in the proposed set(s) of desires. After all, the actual agent's ideal advisors represent the only rationally permissible sets of desires, and strengths of desires, that that agent can have. Any set of desires that differs from one of those sets is not rationally permissible—i.e., it would be *irrational* for the actual agent to adopt that set and strength of desires. But it's just not plausible that an actual agent's reasons, and strengths of reasons, are grounded in a set of desires that it would be *irrational* for her to adopt.

It might be suggested, for example, that the actual agent's reasons are grounded somehow in *all* of the strengths of all of her ideal advisors' desires at once. But no ideal advisor has the complete set of desires, and strength of desires, had by all the agent's ideal advisors at once (and the same is true for the sum or average of the strengths of their desires). So it would be irrational for the actual agent to adopt, at once, all the desires, and strengths of desires, had by all of her ideal advisors. This proposal fails, again, because it just isn't plausible that an agent's reasons, or strengths of reasons, are grounded in a set of desires that it is irrational for her to adopt. And this is so for any proposed set of desires that isn't already had by one of the ideal advisors. So this pushes us back to the thought that the agent's reasons must be grounded in the desires of one, and only one, of her ideal advisors. But that way lies the arbitrariness problem. Thus, either horn of the dilemma is deeply unattractive for idealizing internalists.

5 Conclusion

To sum up: Idealizing internalists assume that, for any actual agent with an irrational psychology, there is one uniquely rational way of correcting that psychology so that it conforms with the standards of structural rationality. They then derive the facts about the actual agent's reasons from this uniquely rational counterpart. But I've argued that the uniqueness assumption is false. There are at least some cases in which there is no uniquely rational way to correct an irrational agent's psychology. And this shouldn't be surprising. The standards of structural rationality typically require only that our attitudes hang together in the correct way, and there are many ways for attitudes to hang together correctly.

The best way to challenge the argument I've been advancing, I think, is to challenge Non-Uniqueness—that is, to argue that, for *any* structurally irrational agent, there is only one correct way to make them rational. But in order to do this,

one needs to defend a *very restrictive* conception of structural rationality. This would certainly help idealizing internalists avoid the difficulties for their view that I've pointed out in the second part of this paper. But it would also give rise to a new problem. The more constraints on structural rationality one posits, the more implausible the account of structural rationality becomes. So there are tradeoffs. Idealizing internalists can either defend a plausible, but less restrictive, account of structural rationality, in which case they'll have to confront the problems for their view associated with Non-Uniqueness. Or they can block any problems associated with Non-Uniqueness, but only by advancing a highly restrictive, and therefore highly implausible, conception of structural rationality. Either way, idealizing internalists confront serious difficulties for their view.

Acknowledgements For helpful feedback or conversation on previous versions of this paper, I owe many thanks to Alex Worsnip, Chris Howard, Alfredo Watkins, Joseph Porter, Omar Fakhri, Robert Reed, Justin Morton, Uriah Kriegel, Kieran Setiya, Bart Streumer, Sarah Stroud, Aliosha Barranco Lopez, Dominic Berger, Daniel Kokotajlo, Anne Jeffrey, Chris Blake-Turner, Alex Campbell, Joanna Lawson, Keshav Singh, Laura Sampson, and two anonymous reviewers.

References

Cuneo, T. (2007). *The normative web*. Oxford: Oxford University Press.
Darwall, S. (1983). *Impartial reason*. Ithaca: Cornell University Press.
Enoch, D. (2011). On mark Schroeder's hypotheticalism: A critical notice on *Slaves of the Passions*. *Philosophical Review, 120*(3), 423–446.
Evers, D. (2014). In defense of proportionalism. *European Journal of Philosophy, 22*(2), 313–320.
Heathwood, C. (2008). The problem of defective desires. *Australasian Journal of Philosophy, 83*(4), 487–504.
Lord, E. (2014). The real symmetry problem(s) for wide-scope accounts of rationality. *Philosophical Studies, 170*, 443–464.
Manne, K. (2014). Internalism about reasons: Sad but true? *Philosophical Studies, 167*, 89–117.
Manne, K. (2016). Democratizing humeanism. In E. Lord & B. Maguire (Eds.), *Weighing reasons* (pp. 123–141).
Markovits, J. (2014). *Moral reason*. Oxford: Oxford University Press.
Parfit, D. (2011). *On what matters* (Vol. 1). Oxford: Oxford University Press.
Rieder, T. (2016). Why I'm still a proportionalist. *Philosophical Studies, 173*(1), 251–270.
Scanlon, T. M. (1998). *What we owe to each other*. Cambridge: Harvard University Press.
Schroeder, M. (2007). *Slaves of the passions*. New York: Oxford University Press.
Shafer-Landau, R. (2003). *Moral realism: A defence*. Oxford: Oxford University Press.
Shafer-Landau, R. (2012). Three problems for Schroeder's hypotheticalism. *Philosophical Studies, 157*, 435–443.
Smith, M. (1994). *The moral problem*. New York: Blackwell.
Sobel, D. (2017). *From valuing to value: Toward a defense of subjectivism*. New York: Oxford University Press.
Street, S. (2008). Constructivism about reasons. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics, volume 3* (pp. 207–245). Oxford: Oxford University Press.
Williams, B. (1981). Internal and external reasons. In *Moral luck* (pp. 101–113). Cambridge: Cambridge University Press.