

Personal ideals and the ideal of rational agency¹

Sarah Buss

Department of Philosophy, University of Michigan

Correspondence

Sarah Buss, Department of Philosophy,
University of Michigan.
Email: sbuss@umich.edu

Abstract

All of us have personal ideals. We are committed to being good (enough) friends, parents, neighbors, teachers, citizens, human beings, and more. In this paper, I examine the thick and thin aspects of these ideals: (i) their *substance* (to internalize an ideal is to endorse a particular way of being) and (ii) their *accountability to reason* (to internalize an ideal is to assume that this is really a good way to be). In considering how these two aspects interact in the ideal of rational agency, I address two philosophical debates that are generally conducted in isolation of each other: (i) debates over the anti-ideal of normative “fetishism” and (ii) debates over whether acting for a reason is acting “under the guise of the good.” In the final two sections of the paper, I further explore the relations among the thick and the thin. I note the role that coherence constraints play in the process whereby

¹This paper is the product of more conversations than I can recall. I would like to thank everyone who has talked with me about the issues it addresses. This includes audiences at The Humboldt University, Tulane, Notre Dame, Duke, the Italian Society of Analytic Philosophers, the Rocky Mountain Philosophy Conference, the Annual Ohio Philosophy Association conference, Joseph Raz's Speakers' Seminar, Heidelberg University, the University of Bern, the CUNY Graduate Center, the University of Texas (El Paso), the University of Toronto, the University of Calgary, Purdue University, the annual meeting of the Northwestern Society for the Theory of Ethics and Politics, Georgia State, UCLA, and the University of Arizona. I am especially grateful for editorial comments Elijah Millgram gave me when I was looking for words to cut. The paper also profited from the feedback Karen Jones and Francois Schroeder gave me on an earlier paper, and from the comments of an anonymous referee at the final stage.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Philosophy and Phenomenological Research* published by Wiley Periodicals LLC on behalf of Philosophy and Phenomenological Research Inc.

our ideals gain determinacy. At the same time, I argue, our ideals constrain the possibility and desirability of coherence. This has implications for a third debate: the debate over the possibility of moral dilemmas.

All of us have personal ideals. We are all committed to being good – good enough – in a variety of ways. I, for example, am committed to being a good enough mother, wife, daughter, sister, philosopher, teacher, neighbor, friend, citizen, person. I am committed to being brave, kind, just, polite – brave, kind, just, polite *enough* not to qualify as a mean, rude, unjust coward. These commitments play an important role in what I take myself to have reason to do. They also determine my apprehension of certain requirements.

This last point is important. If I have internalized the ideal of being a good enough daughter, then I will see the fact that this person is my mother as *a reason* to treat her in certain ways and not others. I will be disposed to see the fact that my mother is ill as a reason to make sure that she is getting good care. But I will also be disposed to see her illness as *requiring* me to make some not-insignificant effort to help her. This reflects the assumption that there are certain things I could do that would count as not living up to my commitment to being a good enough daughter.²

In stressing the connection between (i) internalizing an ideal and (ii) taking oneself to be subject to certain requirements, I do not mean to deny that there are all sorts of things it would be more or less good for me to do, even though I may not have overriding, or even sufficient, reason to do them. And, of course, it is always possible to realize an ideal to a greater degree. Insofar, however, as our personal ideals are inseparable from our commitments, and hence inseparable from our ends, internalizing them is inseparable from apprehending *constraints* as well as *considerations*. Someone who has internalized the ideal of being a good enough mother can perceive the fact that her son desires a brand new bike as *a reason* to buy him a brand new bike without regarding this fact as placing her under any *obligations*. Indeed, she need have no view about how the “weight” of this reason compares to that of other facts – e.g., the fact that she has limited resources for indulging his desires, or the fact that children tend to become insufferable if they come to expect their parents to give in to their every demand, or the fact that other people’s children would be less likely to starve if one spent less money on one’s own children. Nonetheless, insofar as someone is committed to being a good enough mother, she not only assigns positive and negative value to certain actions; she is also disposed to regard certain circumstances as calling for, or requiring, certain responses.

I want to explore the role that our personal ideals play in our disposition to see facts as reasons and requirements; and I want to consider what a proper understanding of this role can teach us about some vexed issues in moral philosophy and the philosophy of action. My chief aim is to illuminate the relation between acting under the guise of substantive personal ideals and being responsive to reasons. In pursuing this aim, I will bring together some philosophical discussions that are typically conducted in isolation of each other. One of my secondary aims is to encourage further inquiry into the relations among the issues addressed in these discussions.

² As Jay Wallace notes, “Our most important projects present themselves to us as making claims on us. As friends, lovers, Oakland A’s fans, and teachers of philosophy, there are certain things that one simply has to do: provide support for the friend during a difficult phase in her life, follow the play-offs, prepare one’s classes, and so on.” (Wallace, 2006a, p. 304)

My own inquiry begins with the widely shared assumption that in order to be a good enough responder-to-reasons, one must be able to relate to one's circumstances under the guise of the sort of thick, substantive, ideals I mention in the opening paragraph. In examining this conception of rational agency, I will do my best to convey the appeal of the ideal that underlies critiques of "normative fetishism." I will then shift my attention to the fact that every ideal – no matter how substantive, or thick – has a thin, aspirational, aspect. Just as substantive ideals play a role in our responses to reasons, so too, the thin ideal of being responsive to reasons is implicit in every substantive ideal. In calling attention to this feature of rational agency, I will be fleshing out and defending the thesis that in acting under the guise of substantive ideals, we are acting "under the guise of the good."

Having identified the thick and thin aspects of our ideals, I will devote the rest of the paper to the relationship between the thick and the thin. I will focus, in particular, on how coherence constraints make it possible for our ideals to become more determinate. And I will examine the limits on coherence that are imposed by the substantive ideals themselves. Even as the content of each substantive ideal partly depends for its contours on the ideal of coherence and our aspiration to be responsive to reasons, it also imposes a constraint on how coherent our normative orientation can be. This, I will argue, is what grounds the possibility of moral dilemmas – and normative dilemmas, more generally. In the paper's final, and longest, section I will argue that normative dilemmas are possible because with each of our commitments to being good in some way, we impose normative requirements on ourselves that are independent of the requirements associated with every other ideal – and independent, too, of the thin ideal of being responsive to reasons. Another secondary aim of this paper is to call attention to the high price we would have to pay in order to be invulnerable to such conflicting requirements.

Before I begin my investigation, I would like to stress two points. First, though I will, for the most part, be discussing ideals that all of us share, or at least endorse, nothing I say rules out the possibility that someone could have the ideal of being a good enough hit woman, or Mafiosa. Someone can "see" facts as reasons in the sense of interest to me here even if she is profoundly mistaken about the normative significance of these facts.³ My interest in our personal ideals is an interest in these normative seemings. With one exception, I will not be considering whether any of our ideals is justified – or, even more fundamentally, what it would take for an ideal to be justified.

The one exception is the ideal of coherence. In what follows I will assume that, all else being equal, it is good to be coherent.⁴ At the same time, I will challenge the assumption that it is, in principle, always possible to become more coherent without paying an unreasonable price. I will argue that this assumption is incompatible with an ideal of rational agency according to which someone is not a good responder-to-reasons unless she responds to reasons under the guise of the sort of substantive ideals I mention in the paper's opening paragraph. I will also point out that internalizing such ideals has an additional, very important, value: it is essential to acquiring and maintaining a substantial identity.

My second preliminary point is that in determining which reasons we see, our ideals are not themselves among these reasons. Though I may sometimes remind myself that I am committed to being a good enough daughter, and that this commitment has certain implications for how

³ In short, unlike Christine Korsgaard, I am not endorsing the thesis that reasons for action are a function of an agent's "practical identity." (See Korsgaard, 1996, p. 101)

⁴ For all I say here, this could simply be because coherence is a symptom of something else that is good – e.g., responsiveness to reasons. For two recent books that discuss this hotly-contested issue, see Kiesewetter (2017) and Loyd (2018).

I treat my mother, my personal ideals are usually constituents of the points of view from which certain responses appear warranted, and even required. Typically, we internalize an ideal (usually without being aware of this fact) because certain things strike us as good and important, and so we want our responses to reflect this fact, and so we want to be disposed to respond this way. The belief that it is admirable to display a certain pattern of responses, and to be disposed to do so, does not usually figure among the things we take into account in deciding how to act. More carefully, though someone may sometimes be motivated by the consideration that “a good X would never do *that*,” this is usually under circumstances in which her relation to the ideal is a less-than-completely-internal one. In the paradigm case, the internalized ideal plays an explanatory role without playing a justificatory role: it explains why certain features of someone’s circumstances appear to her to be reasons to respond as she does.⁵

1 | AN IDEAL OF RATIONAL AGENCY: RESPONDING TO REASONS UNDER THE GUISE OF SUBSTANTIVE IDEALS

It is widely agreed that someone is deficient as a rational agent if, despite doing what she has reason to do because she believes she has sufficient reason to do it, none of the features of her circumstances strikes her as normatively significant in itself. If her sole aim is to do what she has sufficient reason to do, under this description, then there is, it is said, something “fetishistic” about her motivational structure. Rather than being directly (“nonderivatively”) moved by the features or facts of her circumstances – rather than, for example, being directly moved by the plight of the person who appears to be drowning nearby, or the fact that this person is her wife – the normative fetishist is moved by the fact that by responding to these features as she does, she will pass whatever test is relevant to determining what she has reason to do.⁶

But what, exactly, is the ideal relative to which we fall short if we have to “derive” the significance of our actions and circumstances from some very general criterion for determining the normative significance of any and every fact? By stipulation, the normative fetishist cannot *just see* the particular facts to which she is responding *as reasons*. (I.e., she cannot *feel* the normative force of these facts, where this does not require the self-conscious application of the concept ‘reason.’) This is why she needs to appeal to a principle according to which these facts have this significance. To call this person a “fetishist” is to point out that she attributes a value to being guided by a certain principle without having any independent appreciation of the value of any action, or of any substantive end. This principle does not spell out the rationale for normative perceptions she has independent of any reasoning. Rather, it is a normative proposition she must take into account in order to infer that some fact qualifies as a reason to do something. Because the normative fetishist does not directly perceive the normative significance of any substantive

⁵ In effect, then, what I am calling “personal ideals” are similar to what others have called “psychological schemata.” For a recent helpful discussion, which refers to important earlier work, see Walden (2018).

⁶ Michael Smith contrasts the *moral* fetishist with people who “care non-derivatively about honesty, the weal and woe of their children and friends, the well-being of their fellows, people getting what they deserve, justice, equality, and the like, not just one thing: doing what they believe to be right, where this is read *de dicto* and not *de re*.” (Smith, 1994, p 76) The *normative* fetishist stands in the same relation to practical reasons of all sorts as the *moral* fetishist stands in to *moral* reasons. Smith’s claim, and the use to which he puts it in the debate over the metaphysics of moral reality and moral beliefs, has prompted many responses. See, for example, Dreier (2000); Svarvasdottir (1999); Wallace (2006b); Toppinen (2004); and Johnson King (2020).

features of her circumstances, she makes a fetish out of doing “what I have (sufficient) reason to do.”

Whatever the alleged deficiency of this response to reasons, it is not that the one who responds in this way fails to be *directly moved* by certain facts. Her problem is that if she were so moved, she would find her own responses incomprehensible: it would be as if she had done something for no particular reason at all. In avoiding such an alienated relation to her own behavior, she might need to think through her responses more self-consciously and effortfully than is often (though not always) ideal. But this is also not the essence of her handicap. She would be a normative fetishist even if she responded to her circumstances as swiftly and confidently as the rest of us.

The normative fetishist is rather like someone who cannot see two lines as having two different lengths, but needs to measure them both in order to reach this conclusion – with great rapidity, perhaps, and even by exercising some special mental faculty that has no impact on her perceptions. Comparisons with real-life experiential anomalies also suggest themselves. Consider, for example, the difference between someone who has trouble *seeing* various facial expressions as expressions of emotions, but who has learned to draw the proper conclusions. Consider, too, what distinguishes the ordinary experience as of interacting with one’s beloved from the experience of those who are afflicted with Capgras syndrome. Apparently, it is possible for someone to *see that* someone else has all the features of P (and of no one else) without thereby *seeing* this person as P. If, under these circumstances, one nonetheless manages to convince oneself that P is really not an imposter, there will still be something important one cannot see.

Just as most human beings need not appeal to any general principles in order to appreciate that a human being who *looks like* their beloved *really is* their beloved, so too, most human beings rarely need to appeal to any general principles in order to appreciate how they have reason to respond to their circumstances. But if experiencing the familiarity of another human being is mysterious, experiencing a fact as a reason is even more so. Is this the experience as of this fact’s directing, or permitting, or encouraging one to do something? Or is it merely what it is like when one is quite capable of wondering whether one’s responses to one’s circumstances make sense, and yet this question simply does not arise? Is there really a way it feels when one responds to facts *as reasons*? Or is talk of “experiencing” a fact as a reason no less metaphorical than talk of “seeing” reasons as such?⁷

I confess to being unable to understand how we could *respond directly* to our reasons *as such* without *perceiving* them *as such*. (It is perhaps worth recalling here the distinction evoked a few paragraphs earlier between being directly moved by a fact and being directly responsive to its normative significance.) Nonetheless, I am open to the suggestion that talk of what we “perceive” when we respond directly to reasons as such is, at least to a significant extent, metaphorical. What interests me is the phenomenon of responding directly to facts *as reasons*, and the role that our ideals play in making this possible. In trying to clarify what I have in mind, I have spoken of “seeing” reasons as such, and I will continue to speak this way. But, again, I am not committed to a particular account of what this involves *experientially*, and I am even open to being convinced that I am mistaken in thinking that responding to reasons *as such* involves a distinctive experience.

⁷ Joshua Gert has argued that there is no way to make sense of the perceptual metaphors to which many philosophers appeal when they offer accounts of acting for reasons. (Gert, 2004, pp. 191-193) But the mere fact that what one is doing can be explained (by a spectator) as a response to facts which (according to this spectator) warrant this very response does not suffice to ensure that this response is something one does for a reason (nor even that it qualifies as an action). And if there is something it is like to “see a fact as a reason,” then this experience will not be reducible to anything else. For two additional, very different (and very interesting) engagements with this issue, see Fix ([forthcoming](#)) and Schapiro (2021).

If, as I believe, there is a distinctive seeing-a-fact-as-a-reason experience, then this *explains why* the normative fetishist has to think an extra thought about what counts as a reason for what: the extra thought is compensation for the experiential handicap. If, on the other hand, there is no such experience, then seeing a fact as a reason *just is* relating to it in such a way that one need not think an extra thought from which one derives its normative significance. Those who cannot relate to the normative significance of their circumstances in this way must treat as a premise in their reasoning a normative principle that spells out the significance of a second, factual, premise (the fact that this person is drowning – or that this person who is drowning is one’s wife).

Someone who relates to her circumstances in this way stands in a mediated relation to the normative significance of her circumstances and actions. But such mediation comes in degrees – depending on the nature of the normative principle to which the fetishist must appeal. Consider, for example, two very different such principles: “If my wife is drowning, I have reason to try to save her” and: “If in jumping in the water to save my wife, I would be acting according to a principle that I could will to be a universal law, then I have reason to jump in to save her.” If someone cannot *just see* her wife’s drowning *as* a reason to try to save her, then she is also incapable of *just seeing* the *normative significance* of the action-guiding principle according to which one’s wife’s drowning is a reason to try to save her. She cannot *feel* the *normative force* of this principle. And so there is no more to her *belief that* this principle is a reasonable guide to action than the belief that some individual or group or text has declared this to be so. She might have identified some reason for trusting these sources. Or it might be a mystery to her why these people or texts are authorities with respect to what she has reason to do. In either case, she is like someone who obeys an order with no understanding of why the order itself makes sense.

To say that there is something deficient about rational agents for whom this is the usual mode of being in the world is to call attention to the difficulty we have in understanding what they think they are doing. How can someone care about what she has reason to do without caring about anything else? How can she regard *the mere fact that she has a reason* to do something as itself a sufficient reason to do it? How can she take herself to have satisfied her need to discover a sufficient reason for what she does if nothing, in particular, appears to her to count in favor of acting this way?

Such questions call attention to the normative fetishist’s alienation from the normative significance of her circumstances and her action. But someone’s alienation need not be this extreme whenever an action-guiding principle plays a decisive role in her response to her circumstances. Of course, if someone relies for guidance on the Categorical Imperative simply because “this is what my mother (or the good book) says I have reason to do,” then this person also lacks direct access to the normative significance of her actions. But if she feels the force of the imperative – if (perhaps after she reflects on the matter a while) she *just sees* that whether she has reason to do something depends on whether the maxim of her action satisfies the universalizability condition – then relying on the principle does not merely enable her to conclude that certain facts of her circumstances justify a certain response; she *sees these facts as* justifying this response. She sees them this way because she sees them as facts to which she can respond in this way without violating the Categorical Imperative, and because she sees the Categorical Imperative as a constraint on what she has reason to do.

If we are inclined to think that this less extremely alienated person is also deficient in her response to reasons, then this must be because we endorse an ideal of rational agency according to which a rational agent makes certain *substantive* normative and evaluative assumptions in light of which she sees certain facts as reasons. According to this ideal, a

good-enough-responder-to-reasons acts under the guise of more substantive ideals. These ideals enable her to see facts as reasons to do something; and so, they enable her to form beliefs about what she has reason to do without drawing inferences from any general action-guiding principles.

It is important to stress that one can regard this sort of reasons-responsiveness as an ideal without rejecting the value of self-reflection, or the value of being able to offer a principled justification of one's actions. Critics of normative fetishism can readily concede that a responsible rational agent is prepared to call her normative and evaluative perceptions into question, and to review the considerations for and against them. Their point is that though there are many circumstances in which a well-functioning rational agent *could* reflect on the rationale for attributing a certain normative significance to certain facts, and though there are even some circumstances under which she *would* do so, she rarely *needs* to do this in order to discover what she has reason to do. In calling our attention to this fact, these critics remind us that acting for a reason requires being able to appreciate the normative significance of some facts without engaging in any reasoning. Indeed, even the extreme normative fetishist must be disposed to regard the book of action-guiding principles as authoritative without being moved by any distinct reasons for so doing. If she is to make up her mind to do anything, she must nonderivatively perceive consulting the book as the thing she has reason to do.

2 | ACTING UNDER THE GUISE OF IDEALS

A normative fetishist is someone who fails to *just see* certain facts *as reasons* because she lacks any substantive normative ideals. She cannot directly relate to her circumstances under the guise of thick, substantive assumptions regarding what is important and good. If this is a deficiency in her mode of responding to reasons, then this is because to be a good enough responder-to-reasons, it is not enough that (i) one is disposed to being *good enough at responding to reasons, whatever this may turn out to require*, nor that (ii) this disposition is sufficiently strong to determine how one responds to one's circumstances. One's responses must also be shaped by more substantive commitments.

I want now to explore the implications of that “also”: a good enough responder-to-reasons cannot be indifferent to whether she is responsive to reasons. To the contrary, this very thin ideal is implicit in every other ideal, no matter how thick; one cannot internalize a given ideal without taking oneself to have some reason to realize it. If I have internalized the ideal of being a good friend, then I do not merely see my friend's plight as a circumstance to which a good friend would respond by jumping in the water. I see the fact that my friend is drowning *as a reason* to jump in; and it is *as a reason* that this fact moves me to jump in. Seeing a fact as a reason involves treating a relatively determinate – relatively thick – guide to action (e.g., Try to prevent someone from drowning, especially if this someone is your wife) as a stand-in for something very indeterminate, or thin (e.g., Do only what you have good enough reason to do). It is only because the nonfetishist can directly apprehend the relatively thick as a specific instance of the vanishingly thin that she does not need to fetishize acting for reasons as such.

When we do things for reasons, we aspire to achieve certain determinate ends because we think that so doing makes sense; and we think that so doing makes sense because we think that achieving these ends is in some respect good. The “thinness” of this aspiration consists of the fact that

we can never reach a point at which we fully understand what is involved in “living up to our” ideals, and that it is always possible that what appears to be a gain in insight is actually a step in the wrong direction. Regardless of whether the ideal in question is the ideal of motherhood, friendship, or rational agency, we can always intelligibly wonder whether what *appears to be* a good way to be *really is* a good way to be, or whether we have properly understood what it is to be good in this way. In short, every conception of every personal ideal is necessarily to some extent indeterminate and provisional; and this is because every conception of every personal ideal takes the form of an approximation, or stand-in, for something thin: whatever really counts as a good way of being.^{8,9}

*

If in aiming to realize various ideals, our target is necessarily pretty fuzzy – if it may not even be in the place we are seeking it – and if in determining what we have reason to do on any occasion, we rely on a plethora of internalized ideals, then acting for reasons essentially involves being motivated by aspirations we only dimly comprehend. Indeed, insofar as substantive aspirations include the aspiration to be in some respect *good*, there is an important respect in which they incorporate something indeterminate, or thin.

We can put this point by saying that there is a respect in which each substantive aspiration is “empty.” Importantly, however, this is not the respect in which one’s goal is empty insofar as it is simply “whatever I in fact aim at.” We must thus reject David Velleman’s argument against the assumption that acting for a reason is acting under the guise of the good.¹⁰ According to Velleman, “the good” could not be the aim of action for the simple reason that in aiming at “the good,” we would be aiming at nothing at all. “The good,” Velleman says, is an empty concept. Aiming to do what is good is like aiming to capture “the quarry”: unless one has some determinate conception of what the quarry is, one has nothing to hunt (Velleman, 1996, pp. 700–702).¹¹

If the preceding observations are correct, however, then Velleman’s comparison is inapt. The fact that “the good” is “empty” just is the fact that our substantive ends are aspirational in a way that goes beyond the fact that we are disposed to try to achieve them. We have certain substantive ends insofar and only insofar as we take these ends to be *worth* pursuing; “the

⁸ As I hope will be clear, much of the discussion that follows reflects my sympathy with David Wiggins’s claim that “the subject matter of the practical is both indefinite and unforeseeable.” (Wiggins, 1975-76, p. 41) “There does not,” Wiggins elaborates, “exist the option . . . for an agent to measure in advance what exactly any kind of commitment lets him in for, either in and of itself or in relation to all his distinct commitments (whatever *they* may prove to amount to). It cannot be predicted in the real world how much scope one positive commitment will allow to others. Nor can it be predicted what it will take to persist in a given commitment.” (Wiggins, 1998, p. 372)

⁹ Though I am not here directly engaging with the debate over whether “good” has a “predicative” application/use, in defending the aspirational aspect of the ideals under whose guise we act, I am, in effect, arguing that when an ideal is our own, the judgment that something is good in the relevant way is not simply the judgment that it is good *of its* kind. (For the classic argument against the “predicative” use of ‘good,’ see Geach 1956. See also Thomson (1992).)

¹⁰ The critique I will be offering here differs in several important respects from the critique Philip Clark makes in “Velleman’s Autonomism” (Clark, 2001). Perhaps the most fundamental difference is that Clark’s critique rests on the claim that “the good” under whose guise agents act is a substantive, not a formal, end.

¹¹ See especially, pp. 700-702. In raising this complaint, Velleman is anticipated by Murdoch: “The idea of perfection might be, as it were, empty.” (Murdoch, 1991, p. 61) As I indicate in the text that follows, my response to Velleman is, essentially, Murdoch’s.

good” is, essentially, *that which the substantive ends are approximating*. To be sure, we must form particular, determinate conceptions of the good in order to have something to aim at in action. But satisfying this requirement is compatible with regarding these conceptions as imperfect and provisional – as approximations of something, we know not what.¹² Indeed, if what I have said is correct, then to regard such stand-in ideals as something worth realizing just is to take for granted the intelligibility of wondering whether they really are worth realizing.

Just as none of us is a normative fetishist, so too, none of us is entirely unaware of the aspirational aspect of our substantive aims. This means that we generally take it for granted that these aims can be intelligibly challenged. When someone’s conception of what it is to be a good mother seems to imply that she ought to allow her daughter to wear her pajamas to school, she may find it quite natural to ask herself: “But is this really what a good mother would do?” All questions of this sort are *open* questions. Any conception of how it is good to be can be challenged as long as it is a substantive proposal.¹³

No similar challenge makes sense if we have decided to hunt for mushrooms. Precisely because what we hunt is a matter of stipulation, if we know what that stipulation is, it makes no sense to ask: “But is this *really* the quarry? Might it not really be *foxes* instead?” This question makes no sense because the only aspiration involved in hunting for mushrooms is the aspiration of finding ... *mushrooms*.

If our aspiration is to find the *best* mushrooms in the forest (the ones it would make *most sense* to seek), then something “empty” (something “thin”) has crept into our goal. No stipulation can insulate any apparent paradigm case from challenge; of any specimen we discover, it is always intelligible – and sometimes even reasonable – for us to ask: “Does *this* specimen really fill the bill?” So, too, we can intelligibly ask whether hunting for mushrooms is really such a good thing to do. To insist that these questions do not make sense is to be ethically blind. These questions make sense, even though no answers are immune from further questioning. They make sense *precisely because* such closure is not a conceptual possibility.

Again, this is not to deny that we must have ends that are determinate enough to indicate which steps we can and must take to achieve them; we must have ends that can serve as a guide to our choice of means. The point, however, is that insofar as we take ourselves to have reason to achieve these ends, our aim in acting is necessarily also indeterminate enough to represent the distinction between any ideals we might *actually aim* to realize and whatever ideals are really *worth* realizing. It is, in other words, indeterminate enough to represent the distinction between (i) whatever determinate conception we have of what it would take to realize these ideals and (ii) what it would really take to realize them. In acting for reasons, we are guided by determinate-enough ends, conceived as provisional stand-ins for whatever is really and truly the worthy object of aspiration we seek. As Iris Murdoch notes, this is “the true sense of the ‘indefinability’ of the

¹² As Korsgaard notes, “‘Good’ names the problem of what we are to strive for, aim at, and care about in our lives.” ... The ‘thinness’ of the [terms ‘right’ and ‘good’] ... comes from the fact that they are, so far, only concepts, names for whatever it is that solves the problems in question. “We need *conceptions* of the right and the good before we know how to apply the concepts.” (Korsgaard, 2008, p. 322) Without endorsing Murdoch’s realism about the good, Simon Blackburn calls attention to this thin concept’s status as a regulative ideal when he speaks of the process whereby descriptive elements of “thick ethical concepts” are “disentangled” from evaluative elements. According to Blackburn, “approaching a purely descriptive paraphrase of a thick evaluative concept” is like approaching an “asymptote.” (Blackburn, 2013)

¹³ Importantly, to suggest that there is an in-principle-uncloseable gap between what *appears* to be (in some way) good and what *really* is good in this way is to note that the reality in question is of a distinctive sort. (In my view, the asymptotic nature of this reality is essentially tied to the fact that what is real is an infinite regress of supporting reasons. But that is a story for another paper. (See Buss (unpublished manuscript a))

good”: “It lies always beyond, and it is from this beyond that it exercises its *authority*.” “Good is the focus of attention when an intent to be virtuous co-exists (as perhaps it almost always does) with some unclarity of vision.” (Murdoch, 1991, p. 62, 70)

Note that neither “the quarry” nor anything that is stipulated to be the quarry is an end of this sort. “The quarry” (conceived as such) is empty in precisely the sense that any stipulation will count. Precisely for this reason, moreover, no stipulation leaves room for the possibility that the apparently paradigm/exemplary cases are not really what we are looking for after all.

3 | Refining One’s Aspirations: How a Relatively Indeterminate Conception of an Ideal Becomes Somewhat More Determinate

Every conception of a particular ideal is to some extent thick. It consists of a description of what one must do in order to be good enough in the relevant way, and of the sort of responses to various circumstances that are typical of someone whose apprehension of reasons is constrained by the fact that she is good in this way. As our conceptions of our ideals become more determinate, there are more nonnormative facts we can see as reasons without relying on the mediation of additional considerations.¹⁴ But how do our conceptions of our ideals gain determinacy? How do we flesh them out in a way that reflects a more adequate understanding? Under what conditions do *changes in our apprehension* of reasons reflect *greater insight into* the implications of our evaluative commitments?¹⁵

Before I address these questions, I would like to call attention to two distinctions that I will, for the most part, continue to ignore. The first is the distinction between our ideals and our *conceptions* of our ideals. What it takes to be good in certain ways partly depends on our *conceptions* of what it takes. In refining or revising these conceptions, we are thus often altering the ideals themselves. We need not be aware that this is what we are doing. In replacing one ideal with another, it may seem to us that we are gaining deeper insight into certain stable evaluative commitments. In such cases, the connection between our present and former selves will not be as tight as we think.

At the same time, our ideals generally have a determinacy that is not fully captured by our *conceptions* of them. This is, in part, because we inherit many of them from others, and so – like our words (though not nearly to the same extent) – they have a significance that is independent of us. It is also because we can internalize an ideal without being aware of this fact; and because even if we are prepared to avow an ideal, we may not fully understand what the relevant commitment amounts to. For these reasons, changes in our conceptions of our ideals need not correspond to changes in the ideals themselves.

The second distinction I am going to ignore concerns two (compatible) ways in which one can gain a fuller understanding of a given way of being good: (i) by reflecting on one’s commitments and determining what further responses they permit, encourage, require, and forbid; or (ii) by developing the capacity to apprehend an increasing range of facts as reasons and requirements.

¹⁴ Note that as our ideals gain determinacy, we can also come to attribute a *different* normative significance to certain facts.

¹⁵ As I note at the end of this section, this issue is central to discussions of “specificationist” reasoning. As Henry Richardson puts it, the question is: “What licenses us to call a modification or sublimation of an original norm still in some significant sense the same norm that we started out with?” (Richardson, 1990, p. 291) Five years after raising this question, Richardson explains that “it was to the challenge of continuity or stability, above all, that my model of specifying norms was addressed. It helps answer the question: How is it that a norm is being brought to bear on some particular problem even though the interpretation of that norm shifts?” (Richardson, 1995, p.130)

In the first case, gains in understanding one's evaluative commitments take the form of gains in what one can grasp in thought. In the second case, they take the form of a more discriminating set of responses.

Just as a person can internalize an ideal without being aware of this fact, so too, she can acquire a better understanding of this ideal without being able to put what she has learned into words. No matter how self-conscious and articulate an agent may be, moreover, there will always be more to the implicit normative structure of her responses than she can possibly make explicit. No agent or theorist can spell out all the responses that are characteristic of a good *X* under every possible circumstance. Though I will not continue to stress this point and the others I have made in the last few paragraphs, it is important to keep them in mind.

*

When we first internalize our ideals, we have only the dimmest notion of what it takes to realize them. As we begin filling in our conceptions of ways we are committed to being good, we are under pressure to render them more determinate in such a way that they are in harmony with each other. This pressure to achieve and maintain coherence among our commitments is, in part, metaphysical: each of us is only one person; so, however varied our ideals may be, we are under pressure to reconcile our commitments. But it also reflects a normative constraint: if two ends are incompatible, then we seem to have reason to revise (or even abandon) at least one.

It is their appreciation of these two constraints that has led many philosophers to endorse the unity of the virtues. Unless improving our understanding of what it is to be good in some way involves nothing more than learning which means (constitutive and instrumental) are necessary to realizing this good, any gain in understanding will have to depend on our knowledge of what is to be good in *other ways too*. Fleshing out any given ideal will thus necessarily involve incorporating material from independent normative and evaluative assumptions.

How does this work? How does a limited understanding of what it is to be good in some way provide us with a guide to how to rely on *independent* considerations in order to acquire a more complete understanding? I wish I had the answer to this question. I wish I understood how acquiring more determinate conceptions of the various ways we want to be good can be a learning process.¹⁶ Without trying to offer anything approaching a full account of this process, I would like to call attention to three, closely related, factors that seem to be relevant to our ability to gain a better understanding of the constraints we impose on ourselves in internalizing various substantive ideals. First, as I suggested in the previous section, if some ideal *I* really is one of my ideals, then I assume (though, of course, not necessarily self-consciously) that some considerations count in favor of taking up the point of view that is constituted by *I*. This means that any facts that independently appear to me to be reasons for or against various responses to my circumstances are, for me, presumptive constraints on what can count as realizing ideal *I*. If, for example, it seems to me that *under these very circumstances*, I must decline this beggar's request in order to realize ideal *I*, but I am also convinced that there are decisive reasons to give him some money *under*

¹⁶ Though he does not offer the sort of account I am looking for, Talbot Brewer does give a very nice description of what the process is like. In describing a singer's gradual improvements in interpreting a song, he notes that she is guided by her "grasp of a kind of goodness present in an interpretation of the song of which she has an as-yet-indistinct intimation." (Brewer, 2012, p. 47)

these very circumstances, then it seems to me that I have reason to rethink my conception of what is involved in realizing ideal *I*.

I have also already identified a second factor that enables us to gain a deeper understanding of what it is to be good in some way: because we have many different ideals, the aims associated with each one impose limits on what we must do in order to realize the others. If, for example, I have the ideal of being a good mother, then if I also have the ideal of being a kind person, this second ideal will play a role in my understanding of what a good mother would say when her son asks her what she thinks of his art project.

The first of these two factors reflects the normative constraint imposed by the aspiration to be responsive to reasons: it must be possible for us to justify each of our evaluative commitments in light of the others. The second factor also reflects a metaphysical constraint: each of our commitments must be able to make room for the others.¹⁷ A third factor that interacts with these two is the constant alteration of our circumstances – their alteration in many ways we are not able to anticipate when we first internalize our ideals. My incomplete, imperfect understanding of what it is to be a good mother may not spell out what a good mother would do when her daughter insists on wearing her pajamas to school; or it may seem to imply that I ought to do something which, given my other ideals, I am pretty sure I ought *not* to do. In either case, in confronting this novel circumstance, I am forced to reconsider what it is to be a good mother.

In this example, the new circumstance is a new request. It is worth stressing that the changes that force us to rethink our ideals include *social* changes. Thus, to take an example from Carl Elliot, for many centuries someone in China who aspired to be a good parent would bind the feet of her daughter. But as social circumstances changed, “If you love your daughter, don’t go easy on her footbinding” became “profoundly bad advice.” (Elliot, 2003, p. 247) This case also highlights the difficulties we encounter when we try to determine whether our earlier conception of an ideal was incomplete or whether it was simply confused. Did parents in China merely discover that what was appropriate to do under earlier circumstances is not appropriate under very different circumstances? Or, given that the relevant change in circumstances included (among other things) the development of a more egalitarian social structure (with the accompanying belief that the previous structure was unjust), did they discover that the parents of generations past had not really been the good mothers and fathers they had taken themselves to be?

In stressing the important role that the ideal of coherence plays in our refinement of our ideals, I am endorsing a theme that is prominent in the literature on “specificationist” reasoning. Philosophers interested in the possibility of reasoning about ends have pointed out that, given the indeterminacy of our different concerns and goals, the meta-ideal of coherence provides us with a basis for favoring certain interpretations of these concerns and goals over others.¹⁸ Given, however, that not just anything can count as being a good mother, philosopher, ballet dancer, or lion tamer, there is also reason to be skeptical about the possibility of affecting a perfect reconciliation. The point is not simply that we are finite, mortal beings who cannot play arpeggios while juggling five balls on a high wire. The more interesting, and deeper, reason why so many of our ideals force us to give up the ideal of living up to them all is because they are not *entirely* indeterminate. To be sure, we must be prepared to change our minds about what it takes to live up to our evaluative commitments. But there are limits. Every revision must respect the constraints imposed by the determinate content of the ideals.

¹⁷ For a discussion of the extent to which the pressure to maintain coherence is metaphysical, see Buss ([unpublished manuscript b](#)).

¹⁸ See note 15.

In short, even if we are never in a position to rule out the possibility that we have something to learn about what it is to be good in some way, not any old thing can count as being good in this way. This is closely related to my point in contrasting “the good” with “the quarry”: not only is it impossible for someone to simply decide what will count for her as a good way of being; it is also not possible for her to make it the case that a given action is consistent with her ideals whenever she thinks she has overriding reason to act this way. Even though we have good reason to reconcile our ideals with each other, any such reconciliation requires that each ideal has enough content to do the reconciling. This means that no reconciliation is possible if there are no constraints on which reconciliations are possible.¹⁹

If there is more to a person than her identity as an appraiser of and responder to reasons – if she does not have a merely fetishistic commitment to doing whatever-I-have-sufficient-reason-to-do – then there are likely to be cases in which it seems to her that she cannot respond to the relative weight of her reasons for action without betraying at least one of her ideals.²⁰ What should she do in such cases? Perhaps she should give up one of her apparently conflicting commitments – or replace it with something less demanding in the relevant respect. But this might not be possible: after all, these are *her* ideals; she really is committed to being good in these ways. Even if, moreover, she could make the necessary adjustments, she might reasonably object to doing so. She might reasonably prefer to acknowledge the force of certain requirements at the expense of improving the coherence of her commitments.²¹ If some of these requirements are of the moral variety, this would amount to a preference for being vulnerable to finding herself in circumstances in which she is required to do what she is morally obligated *not* to do. It would, in short, be a preference for being vulnerable to moral dilemmas.

4 | MORAL DILEMMAS AND THE THICK ASPECT OF IDEALS: THE POSSIBILITY OF BEING SUBJECT TO REQUIREMENTS ONE HAS OVERRIDING REASON TO VIOLATE

The possibility of moral dilemmas is the possibility that some apparently irresolvable conflicts among moral requirements really cannot be resolved – that there is no way to eliminate these

¹⁹ The point I am making here is closely related to the point Selim Berker makes in a very interesting critique of particularism. If, Berker argues, “reasons for action are irreducibly context-dependent” (as the particularists claim), then they cannot play the role the particularist assigns them of being “the ground floor normative units undergirding all other normative properties and relations.” If, he explains, we “posit too much irreducible context-dependence in the behavior of reasons, . . . the reason-based framework breaks down.” So too, if we allow for too much reconciliation among our ideals, the ideal-based framework breaks down. (Berker, 2007, p. 109) (Indeed, insofar as the ideal-based framework determines which reasons one sees, it just is a breakdown in the reasons-based framework.)

²⁰ Harry Frankfurt has evoked the notion of “volitional necessity” in stressing the extent to which essential aspects of our identity can conflict with our identity as appraisers of and responders to reasons. (Frankfurt, 1988) Whereas Raz (1990) and McDowell (1979), and others argue that a virtuous agent would “exclude” certain considerations from her deliberations, Frankfurt’s focus is on cases in which someone acknowledges the decisive normative significance of a given consideration, and yet cannot reconcile this verdict with his deepest cares and commitments.

²¹ My position in this paragraph (and in the section that follows) is in the spirit of Robert Adams’s observations about the virtues: “The integration of values that Virtue demands . . . insists (though not impatiently) on consistency, but also seeks persistently to relate one’s core values to what does not seem to make a neat package of them. There is always plenty of such recalcitrant material in our lives, and some of it may have a powerful moral claim on our attention. It is even more important, after all, that our aims be good than that they be integrated. Aspiration for Virtue is not just for conformity of the self with itself, essential as that is.” (Adams, 2006, p. 209)

conflicts by introducing further exceptions, qualifications, and other refinements. If moral dilemmas are possible, then it is possible for an ideal moral agent to rightly believe that she is violating a moral requirement, even as she also rightly believes that she has no more justifiable alternative.²²

I want to explain this possibility by elaborating on the observations in the last paragraphs of the preceding section. To this end, I want to consider circumstances in which a person takes herself to have an overriding reason to do something even though she believes this will involve violating a requirement not to behave this way, under these circumstances. This sort of situation differs from those that are central to most philosophical discussions of “moral dilemmas.” But if requirements can retain their force in the face of *sufficient* reason to violate them, then this is because they can retain their force in the face of *overriding* reason to violate them.

I begin with a reminder of how our substantive ideals render us vulnerable to the experience as of facing a moral dilemma. Imagine that someone has internalized the ideal of being kind. The content of the corresponding commitment is sure to be rather indeterminate. Still, not just any sort of behavior counts as being kind. We can reasonably assume, for example, that living up to this ideal is not compatible with telling a small child how disgusting and stupid and boring and ugly and useless she is, calling her names, taking her prized possessions, locking her in a small dark room, and ignoring her every appeal for help and affection. Yet, surely, one is morally permitted – even obligated – to behave this way if doing so is necessary to save the child’s life. Surely, this is just the conclusion someone will draw if she has an adequate understanding of how her commitment to being kind relates to her other commitments.

Those who reject the possibility of moral dilemmas are committed to one of two interpretations of this situation: they must say that, under these special circumstances, this child-tormentor is not really behaving unkindly; or they must say that, though she is behaving unkindly, it does not follow that she is doing anything she ought not to do. On both interpretations, though the person’s behavior certainly has features that *normally, generally, typically* are normatively (and, in particular, morally) problematic, this is the full extent of their moral significance in this particular case. Since, moreover, her behavior has no further moral significance, she has no reason to regard herself as falling morally short. She may regret that she had to harm someone. But she was justified in doing so – *morally* justified. So, she has no more grounds for regret than she has in any other case in which trade-offs must be made. That she takes herself to have violated a requirement – and even a *moral* requirement – may be a commendable indication of how deeply committed she is to avoiding wrongdoing. If so, it is a commendable mistake.

The obvious motivation for interpreting the situation in this way is that there is something incoherent about acknowledging that one is required to do something on this very occasion even though one believes that one has overriding reason not to do it. Our question, however, is whether we always have good reason to attribute this incoherence to a normative mistake.

This question takes us back to two features of our ideals to which I have already called attention. First, insofar as we have internalized a substantive ideal like the ideal of refraining from treating people cruelly, we not only acknowledge that behaving this way is in some respect good; we also take ourselves to be subject to certain requirements – the requirements we must satisfy in order to avoid behaving cruelly. Second, as I indicated at the end of the paper’s previous section, since each substantive ideal is not entirely indeterminate, the requirements that spell out what we must do in order to realize these ideals cannot be revised in just any way so as to reconcile them to each other. In exploring the relevance of these two points to the possibility of moral dilemmas, I aim to show that, like (i) debates over which thoughts are “too many” when we are determining what

²²No characterization of moral dilemmas is uncontroversial. For a collection of essays on the topic, see Mason (1996).

to do and (ii) debates over the relationship between doing something for a reason and acting under the guise of the good, the debate over the possibility of moral dilemmas is, ultimately, a debate over the relation between the thick and the thin. Moral dilemmas are possible, I will argue, because there is more to the substantive identity of each substantive ideal than its identity as an ideal. Given this fact, we cannot insist that apparent moral dilemmas are *merely* apparent without thereby endorsing an ideal of rational agency according to which being a good enough responder-to-reasons requires manifesting a measure of normative fetishism which is less than ideal.

If, unlike the normative fetishist, our normative perceptions of our circumstances are conditioned by our substantive ideals, then, to this extent, our normative perceptions are not a function of our independent assessment of the reasons for and against endorsing them. To note that our substantive ideals function as a priori conditions for the possibility of seeing facts as reasons and requirements is to note that when they function in this way, this does not reflect our assessment of the reasons for and against their playing this role. From this it follows that there could, in principle, be circumstances under which we reasonably conclude that we have overriding reason to refrain from doing something, even though, by our own lights, acting this way is one of the things we are required to do. In short, the requirements we impose on ourselves by undertaking commitments (setting ourselves ends, internalizing ideals) do not have merely *pro tanto* force. Given their independence from our all-things-considered normative judgments, we do not (and cannot) cease to regard them as requirements whenever we take ourselves to have insufficient reason to comply with them.²³

This point is closely related to another: even as the determinate content of a substantive ideal is a partial, provisional stand-in for “a way we have good reason to be,” it also provides us with the material we need in order to evaluate the plausibility of any normative claims. Even as the aspirational (thin) aspect of our substantive ideals renders them open to external assessments regarding whether they really single out ways of being good, their determinate (thick) aspect imposes constraints on the relevance of external considerations. On the one hand, we cannot understand our commitment to refraining from behaving cruelly without relying on other evaluative beliefs, such as our beliefs about what it takes to treat people respectfully and fairly, and on nonevaluative beliefs, such as our beliefs about which events set back people’s significant interests or interfere with their capacity to set and pursue their own ends. On the other hand, when we internalize the anti-cruelty ideal, we acquire an independent evaluative standard which helps us to determine what we must do in order to treat people respectfully and fairly, and more specifically, what we must do in order to avoid setting back their interests and interfering with their ends in morally unacceptable ways.²⁴

It is this “other hand” that explains why those of us who have internalized the anti-cruelty ideal would almost surely take ourselves to be required to do something that, all things considered, we ought not to do if we found ourselves in the unfortunate position of the woman mentioned above. So, too, if we had two children, and if we had to torture one of them in order to save the other, the mere fact that it seemed to us that we had sufficient, and even overriding, reason to do this would not prevent us from concluding that, under these circumstances, there is nothing we can do to avoid falling short as a parent. This conclusion would reflect our assumption that whether

²³ Bernard Williams defends this position in “Ethical Consistency.” (Williams, 1965) More recently, Lisa Tessman has defended the same point. (Tessman, 2015)

²⁴ For more on the relation between the thick and the thin, see Hurley (2002); Smith, 2013; Elstein & Hurka (2009); and Raz (2000).

someone violates a duty of parenthood when she tortures her child does not depend on whether, under the circumstances, she really has no morally better alternative.²⁵

It seems, then, that according to our own ideals of good-enough parenthood and cruelty-avoidance, it is possible to do something we ought not to do, even while doing what, all things considered, we have sufficient (and even overriding) reason to do. In short, our own substantive commitments commit us to the possibility of moral dilemmas. Another way to put this point is to say that insofar as our ideals have some substance, they do not form a perfect unity. Again, the point is not just that we have many different incommensurable values.²⁶ And so, it is not just that in giving some considerations priority over others, “something very significant will [often] be irretrievably lost” (Railton, 1996, p. 148). Though we may reasonably feel regret at failing to realize and express our appreciation of one or more values, my concern here is with the regret and contrition we reasonably feel in response to the fact that we have violated (indeed, defied) an acknowledged requirement.²⁷

*

It is worth stressing that our ideals would not form a perfect unity even if the aspiration to such unity were an essential aspect of each of our substantive ideals. Consider, for example, someone whose commitment to behaving kindly is a commitment to being kind-in-such-a-way-that-I-can-also-be-just.²⁸ Even under these circumstances, the unity of her virtues cannot be a *perfect* unity – as long as there is a more substantive aspect to each of these virtues than the disposition to do (and feel)-what-I-have-sufficient-reason-to-do (and feel). According to the person we are here considering, what it is to be kind depends in part on what it is to be just, and what it is to be just depends in part on what it is to be kind. But this means that, on her view, what it is to be kind and just is a function of normative constraints that cannot be reduced to the requirement to respond to reasons. In short, if, as Aristotle suggests, acquiring practical wisdom involves refining various “natural” virtues that do not themselves presuppose such wisdom (Aristotle, 1941, Book VI, Chap.

²⁵ These observations are relevant to discussions of “thick ethical concepts.” The concepts with which we pick out our personal ideals have both a descriptive and evaluative content. Though for any nonevaluative characterization of an action, it is an open question whether it is good for someone to act this way, certain ways of describing an action are incompatible with its being good in the way that a *kind* action is good. Actions that can be accurately described in these ways cannot be kind. This is something we know as long as we know what we mean by “kind.” It is in this sense that the evaluative component of the concept is inseparable from the descriptive component: to be kind just is to be good in a certain way; so, if no good actions can have the descriptive features of kind actions, then the concept has no application; there can be no kind actions. For some recent work on thick ethical concepts, see Väyrynen (2013) and Kirchin (2017).

²⁶ For a collection of essays on the role that incommensurable and/or incompatible values play in our choices, see Chang (1997).

²⁷ It is, I think, important to distinguish cases of self-betrayal from cases in which one’s action is easily (mis)interpreted as reflecting the failure to endorse an ideal. In advocating a broader conception of “moral dilemmas,” Railton notes that insofar as our actions appear to be incompatible with some of our most deeply held values, they “‘taint’ or ‘compromise’ us in ways we find destructive of the messages we ordinarily hope to convey with our actions.” (Railton, 1996) Clearly, this can be a significant cost – and this cost can be a rational object of regret. It nonetheless differs from the costs of betraying a commitment. Notice that if I am right in suggesting that such betrayal warrants feelings of guilt, then moral dilemmas represent a challenge to the assumption that there is a perfect symmetry between guilt and indignation, such that whenever I am justified in reproaching myself, others are justified in regarding me as blameworthy.

²⁸ For a discussion of how such interdependent ideals can ground a bottom-up form of moral rationalism, see Buss (2018).

13, 114b1-1145a12, pp. 1035–36), then the necessary conditions of practical wisdom are limits on the possibility of the unity of the virtues.

Of course, even if a champion of the anti-dilemma ideal of rational agency were to concede that each of the ways we are committed to being good has its own distinct identity, and even if she were to concede that such substantive ideals are a necessary condition for the possibility of our directly perceiving the normative significance of a wide range of facts – she might still insist that the experience of facing a moral dilemma is necessarily illusory. She might insist that whenever the conceptually necessary less-than-perfect unity of our ideals manifests itself in an apparent dilemma, we have sufficient reason to conclude that (i) we are mistaken about what at least one of our commitments commits us to doing or (ii) we need to modify at least one of these commitments so that it no longer seems to us that we are required to do what we lack sufficient reason to do. According to this proposal, though we can never be fully coherent, we have reason to get as close as possible by making the adjustments necessary to avoiding normative dilemmas whenever it seems to us that we face them. Indeed, according to this proposal, making such adjustments is what we are rationally *required* to do: if we do not make these adjustments, we will not be coherent enough to be good enough responders to reasons.

The problem with this suggestion is that it ignores the price of following it: alienation from one's own substantive commitments, or – if this is different – the absence of any such commitments. This charge should sound familiar. According to Bernard Williams, it is just this sort of self-alienation that necessarily characterizes the conscientious act utilitarian; and this is in large part because the conscientious act utilitarian has no ends that cannot be reduced to the single end of doing what she has most reason to do (where this depends, in turn, on the relative value of the various alternatives). (Williams & Smart, 1973) Rather than exploring the relation between Williams's critique and mine (and between his critique of utilitarianism and his own defense of the possibility of moral dilemmas (Williams, 1965)), I will press the concern he and I share by relating it to my discussion of the normative fetishist.

Consider someone who rejects the possibility of normative dilemmas. Imagine that, like us, this person has many heterogeneous substantive ideals. This means that there will be circumstances under which she will have the experience as of facing a normative dilemma. But when this happens, she will believe that she is confused; she will believe that she is *really* required to comply with only one of the two *apparent* requirements. She will believe this, moreover, even while acknowledging that, as far as she can tell, this disjunctive requirement does not make sense in light of her own substantive commitments, and even though she is thus unable to see this requirement as spelling out, or in any way clarifying, any of her substantive ideals.

What would it be like to relate to the normative appearances in this way? Someone in the situation we are here imagining would, it seems, have no choice but to include among the premises in her reasoning the proposition that anyone who cannot avoid doing A or B is morally required to do only A or B.²⁹ She would have to add this permission to her premises because, though she would *appreciate that* she has sufficient reason not to do A or not to do B, she could not help *seeing* each omission as a violation of a requirement. Given the structure of her normative point of view, she could not help experiencing herself as facing a moral dilemma. And so she would have to think an extra thought from which she could derive the anti-dilemma belief; she would have to relate to this normative aspect of her circumstances in the manner of a normative

²⁹ This position has been defended by several philosophers. As Horty (2003) notes, according to the “disjunctive account,” “the correct all things considered conclusion to draw in these situations is, not that the agent ought to perform the action A and ought also to perform the action B, but simply that the agent ought to perform either A or B.”

fetishist. Worse still, this would not save her from incoherence. For she would continue to see her circumstances as requiring a certain response, even while believing that there is no such requirement.

From what I can tell, a rational agent could maintain this sort of alienation from her own experience of normative reality. But most of us would find it very difficult to do so. More importantly, this difficulty is both the cause and the effect of our assumption that an ideal rational agent would be open to acknowledging competing requirements.

*

A person needs *reasons* to revise her conception of what is required in order to be kind or modest or a good enough mother. And not just any reasons will do. In particular, and trivially, if neither of the ideals that underlie two conflicting requirements provides a justification for altering either requirement in a way that eliminates the conflict between them on some occasion, then this fact is not itself a reason to alter either one of the requirements (or one's interpretation of them). To be sure, if someone has any ideals whatsoever, then she has the meta-ideal of being able to realize them all; and so she has the meta-ideal of being able to reconcile them with each other. But from this it does not follow that nothing short of a full reconciliation is good enough. Indeed, it does not follow that it is better for someone's ideals to be in perfect harmony with each other than it is for there to be some irresolvable tensions in her normative commitments.

Of course, someone who rejects the possibility of normative dilemmas could agree that the condition I have described is far from ideal. She could, however, insist that the problem lies with the fact that the person we are imagining is committed to being good (enough) in many different, incommensurable, ways. Given that the normative sensibility of at least most human beings is at least partially shaped by just such a set of heterogeneous ideals, this would be a significant concession on the part of the champion of the anti-dilemma ideal. She would, in effect, be conceding that her ideal of rational agency is an *inhuman* ideal. More importantly, once we see how significantly human beings would have to be transformed in order to be invulnerable to experiencing normative dilemmas, we are forced to conclude that this would be an even worse way for any rational agent to be.

Trivially, if someone's normative take on the world were not shaped by any commitments whose substance was partly captured by normative requirements, then she would have no inclination to think she was subject to requirements that are at odds with her commitment to doing what she has most reason to do, all things considered. But such a person would secure her invulnerability to normative dilemmas at the price of a pitifully attenuated identity.³⁰ Though she might well appreciate a wide range of values, and though she might thus be disposed to see a wide range of facts as reasons, she, too, would resemble the normative fetishist in being unable to directly see her circumstances as calling for a certain response. She would have to infer the proper response

³⁰ The point I will press here is in the spirit of Elijah Millgram's observation that some "ends cannot be given up or greatly modified without giving up on one's person or character. . . . [T]he price of the [revisions necessary to have a perfectly consistent preference set] would amount to deleting core components of one's personality." (Millgram, 2002, pp. 217-18) Korsgaard (1996, p. 102) has made a similar point: "To violate [the conceptions of ourselves that are most important to us]," she argues, "is to lose your integrity and so your identity; and to no longer be who you are." (Korsgaard, 1996, p. 102). See also Rawls (1999, p. 554): "Although to subordinate all our aims to one end does not strictly speaking violate the principle of rational choice . . . it still strikes us as irrational, or more likely as mad. The self is disfigured and put in the service of one of its ends of the sake of system."

to her circumstances in every case; it would have to be a conclusion she reaches on the basis of weighing various pros and cons. More importantly, though she could regret having to forego certain goods in order to do what, she thinks, she has most reason to do, she would lack the sort of substantive identity that would make it possible for her to betray some aspect of herself without betraying her commitment to being responsive to reasons. In short, there would be no self for her to betray other than the self that is constituted by this very thin commitment.³¹ Her capacity for *self*-betrayal would be limited to her capacity to betray her *circumstance-relative normative judgments*.³² Accordingly, nothing could qualify for her as a case of self-betrayal unless clear-eyed akrasia were possible.

To appreciate this point we need merely consider what would happen when such a person reached a verdict regarding what she has overriding reason to do. If before she reviewed the reasons for and against various possible actions, there had been something, in particular, she thought she stood for, and if she now believed that this substantive commitment was incompatible with what she has overriding reason to do, then she would conclude that she was mistaken about what she stood for – or (though this verdict is, for all practical purposes, equivalent) that, because her previous commitment has no independent normative force, she must revise it into conformity with her all-things-considered normative judgment, or abandon it entirely. Though it might for a moment seem to her as if she was confronting a normative conflict, this would be a mere fleeting impression. In being struck by this appearance, she would be prompted to recall that her identity is restricted to that of an appraiser of actions and circumstances, someone who values certain things and not others, but who stands for nothing more substantive than complying-with-my-*verdicts-regarding-what-I-have-sufficient-reason-to-do-given-how-the-values-add-up*.

As almost always occurs when one engages in debates about fundamental philosophical issues, it is difficult to make the case for or against the possibility of normative dilemmas without appealing to considerations that are just as controversial as the thesis one is defending. Indeed, the significance of these considerations is often controversial for the very same reasons. (Hasn't each disputant just begged the question against the other?) Having followed the argument of the last few pages, the skeptic about normative dilemmas might be inclined to insist that I have jumped to conclusions in assessing the costs involved in being willing to adjust one's (understanding of one's) ideals whenever it seems one is facing conflicting requirements. In particular, this critic might resist my suggestion that the substantive aspect of our ideals – their determinate content – prevents us from revising them in ways that are arbitrary *from the point of view of the ideals themselves*. Isn't it possible, she might wonder, for a person to have a genuine substantive commitment, even if this commitment has "unlimited plasticity"?³³ Have I really said anything that would convince someone who did not share my conception of commitments?

The focus of this challenge is the adjective "unlimited." I have already stressed the significant extent to which we can and should adjust our ideals (our conceptions of what they amount to) in response to new circumstances in which this is the only way to avoid facing conflicting requirements. It is also worth adding that, within the constraints of any given ideal, there may

³¹ Interestingly, in opposing the possibility of moral dilemmas, Barbara Herman implicitly identifies (i) being invulnerable to such dilemmas with (ii) being a person all of whose commitments are subordinated to the single goal of responding to reasons. In arguing that there are no conflicting moral requirements, she approvingly notes that the "central notion" of "the Kantian account of morality" is "the agent as rational deliberator." (Herman, 1996, p. 179) This is precisely the conception of rational agency I am here characterizing as impoverished.

³² Note what follows: if clear-eyed akrasia is not a conceptual possibility, then there is no conceptually possible form of witting self-betrayal or self-defiance.

³³ This challenge comes from a referee for this journal.

well be no limit to the number of possible variations on the theme of how to live up to it. This having been said, just as the fact (which I am prepared to concede, at least for the purposes of this discussion) that (i) there is no end to the possible ways I can sit on my chair is compatible with the fact that (ii) certain postures (including many that are *physically* possible) just do not count as sitting, or (more specifically) chair-sitting, or (more specifically still) sitting on this chair, so too, the fact that (i) there is no end to the possible ways I might interpret my commitment to being a (good enough) mother is compatible with the fact that (ii) certain responses to my circumstances just do not count as living up to this commitment. Again, when I confront such limits, I could choose to adjust my commitment. But, again, from the fact that this adjustment would be necessary in order to avoid facing conflicting requirements, it does not follow that the benefit of greater coherence outweighs the costs. Minimally, there is the cost of abandoning a commitment – or (if this is different) the cost of abandoning my understanding of what this commitment amounts to. But if – as I am imagining – this commitment (this self-understanding) is an essential aspect of one's identity, then to *give it up* is to *give up on* (an important aspect of) *oneself*. Perhaps even more significantly, insofar as one is prepared to make this sort of adjustment whenever this is necessary to avoid facing conflicting requirements, there is an important sense in which one is alienated from one's own commitments. (“Yes, this determinate commitment is an essential aspect of my identity. But it is not *really* essential, after all.”)

Insofar as the skeptic about normative dilemmas is not persuaded by these considerations, it is because she thinks not only that our commitments *need not* have a determinacy that is independent of our assessment of our reasons, but also that they *cannot* have any such determinacy. If nothing I have said constitutes a decisive argument against this alternative conception of commitments, I nonetheless hope that in exploring the conception of rational agency as acting under the guise of substantive ideals, I have shed light on what is at stake in the debate between the skeptic and me. Even if it were psychologically, not to mention conceptually, possible for someone to internalize only those ideals of which she was certain that she could acquire a thicker conception without generating any conflicts with her other ideals – or, if this is different, even if someone could internalize substantive ideals she was disposed to abandon or reinterpret whenever this was the only way to render them compatible with her other commitments – would anyone really have reason to aspire to such a condition? It seems to me that the burden is on the opponent of normative dilemmas to show that her ideal of rational agency is attainable, and if attainable, desirable. Absent such a demonstration, we have good reason to doubt that this ideal corresponds to our own conception of what is required to be a good (enough) rational agent.³⁴

*

Since each of us is a single person, each of us can live only one life. Each of us can only walk in one direction at a time. This means that we are under pressure to unify our ideals: to the extent

³⁴ It is worth noting that this argument challenges an assumption that is central to Ruth Marcus's influential paper on moral dilemmas. Moral dilemmas are, indeed, possible, Marcus argues. But we have reason to “arrange our lives and institutions with a view to avoiding such conflicts.” (Marcus, 1980, p. 121) My point is that just as we can be too risk-averse for our own good in our efforts to avoid false beliefs, so too, there are likely to be significant costs to being strongly averse to making incoherent commitments. As Marcus suggests, we should do what we can to remake our world in ways that make us less vulnerable to the sort of conflicts I have discussed in this paper. But if I am right, the source of these conflicts is not merely circumstantial, or contingent on factors that could possibly be different *for us*. People will always be vulnerable to the sorts of conflicts I have been exploring as long as they have many irreducibly heterogeneous commitments; and this is because no person can live more than one life.

that our ideals do not form a perfect unity, we cannot make sense of ourselves as the subject of these ideals.

This metaphysical pressure to be coherent is, we have seen, reinforced by the aspirational nature of every ideal. Each of our ideals concerns a way of being *good* – or so we believe. We assume that there are compelling reasons for us to realize each of these ideals – compelling reasons to develop certain character traits and adopt certain roles. If, moreover, we are justified in internalizing all these ideals, then this gives us a reason to believe that they can be reconciled with each other. So, if it seems that this is not really possible, we are under normative pressure to reinterpret, or revise, at least one of them.

But this is only half the story. For, as we have seen, there are counter-pressures too. If, like the rest of us, a person has a heterogeneous set of ideals, and if she has internalized these ideals with only the dimmest understanding of what it is to be good in the relevant ways, then she could not achieve full coherence without abandoning many of her deepest commitments. But to abandon one's commitments (or even to be committed to doing so) in order to inoculate oneself against any normatively problematic psychic conflict would greatly impoverish one's life; it would cheapen these commitments to such a point that one would not really be thus committed. Even if, moreover, someone who thus lacked a substantial self would be more coherent in a certain respect, she would also be *less* coherent in another respect. To put the point somewhat paradoxically: if one were committed to altering one's substantive commitments whenever so doing proved necessary to preserve one's coherence, then each of these commitments would itself be less-than-fully coherent.

Whether we have reason to be incoherent in certain ways depends on whether this is a better way of being than any of the alternatives. Toward the end of his novel *The Cave*, Jose Saramago tells us that “a person or character contradicts himself within the bounds of his own inner coherence” (Saramago, 2003, p. 181). This, Saramago implies, is just as it should be. The present inquiry suggests that he may well be right. I have tried to show that there is something more to (at least our kind of) rational agents than a collection of more or less well-integrated dispositions to value some things more than others, to appraise reasons, and to be moved by our appraisals. This means, I have tried to show, that even as coherence pressures play a crucial role in determining our identity, we cannot be who we are without being vulnerable to incompatible (contradictory) requirements. I have argued that no alternative way of being is compatible with the ideal of rational agency that underlies the critique of normative fetishism. The only way to avoid such vulnerability, I have argued, would be to grant the thin aspect of our ideals absolute priority over the thick, more substantial, aspects. Do we really believe that having an identity this hollowed-out is a good way to be?

REFERENCES

- Adams, R. (2006). *A Theory of Virtue: Excellence in Being for the Good*. Oxford: Oxford University Press.
- Aristotle. (1941). *Nicomachean Ethics*. Ross, W.D. (Trans.). In McKeon, R (Ed.), *The Basic Works of Aristotle* (pp. 927–1112).
- Berker, S. (2007). Particular Reasons. *Ethics*, 118, 109–39. <https://doi.org/10.1086/521586>
- Blackburn, S. (2013). Disentangling Disentangling. In Kirchin, S. (Ed.), *Thick Concepts* (pp. 121–36). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199672349.003.0007>
- Brewer, T. (2012). *The Retrieval of Ethics*. Oxford: Oxford University Press.
- Buss, S. (unpublished manuscript a). Against the Quest for the Source of Normativity.
- Buss, S. (unpublished manuscript b). Norms of Irrationality and the Superficial Unity of the Mind.
- Buss, S. (2018). Moral Requirements and Permissions and the Requirements and Permissions of Reasons. In K. Jones & F. Schroeder (Eds.) *The Many Moral Rationalisms* (pp. 110–44). Oxford: Oxford University Press.

- Chang, R. (Ed.). (1997). *Incommensurability, Incompatibility, and Practical Reason*. Cambridge, Mass.: Harvard University Press.
- Clark, P. (2001). Velleman's Autonomism *Ethics*, 111, 580–93.
- Dreier, J. (2000). Dispositions and Fetishes: Externalist Models of Moral Motivation. *Philosophy and Phenomenological Research*, 61, 619–38. <https://doi.org/10.2307/2653615>
- Elliot, C. (2003). *Better Than Well: American Medicine Meets the American Dream*. New York: W.W. Norton & Company.
- Fix, J. (forthcoming) Intellectual Isolation. *Mind*, 127, 491–520.
- Frankfurt, H. (1988). The Importance of What We Care About. In *The Importance of What We Care About* (pp. 80–94). Cambridge: Cambridge University Press.
- Geach, P.T. (1956). Good and Evil *Analysis* 17: 33–42. <https://doi.org/10.1093/analys/17.2.33>
- Gert, J. (2004). *Brute Rationality: Normativity and Human Action*. Cambridge: Cambridge University Press.
- Herman, B. (1996). Obligation and Performance. In *The Practice of Moral Judgment* (pp. 159–83). Cambridge, Mass.: Harvard University Press.
- Horty, J.F. (2003). Reasoning with Moral Conflicts. *Nous*, 37, 557–605.
- Hurley, S. (2002). Roemer on Responsibility and Equality. *Law and Philosophy* 21, 39–64. <https://doi.org/10.2307/3505121>
- Elstein, D. & Hurka, T. (2009). From Thick to Thin: Two Moral Reduction Plans. *Canadian Journal of Philosophy*, 39, 515–35. <https://doi.org/10.1353/cjp.0.0063>
- Johnson King, Z. (2020). Praiseworthy Motivations. *Nous*, 54, 408–30.
- Kiesewetter, B. (2017). *The Normativity of Rationality*. Oxford: Oxford University Press.
- Kirchin, S. (2017) (Ed.) *Thick Evaluation*. Oxford: Oxford University Press. doi: <https://doi.org/10.1093/mind/fzy055>
- Korsgaard, C. (1996). *Sources of Normativity*. Cambridge: Cambridge University Press.
- Korsgaard, C. (2008). Realism and Constructivism in Twentieth Century Moral Philosophy. In *The Constitution of Agency: Essays on Practical Reason and Moral Psychology* (pp. 302–326). Oxford: Oxford University Press.
- Loyd, E. (2018). *The Importance of Being Rational*. Oxford: Oxford University Press.
- Marcus, R. (1980). Moral Dilemmas and Consistency. *Journal of Philosophy*, 77, 121–36.
- Mason, H.E. (Ed.) (1996). *Moral Dilemmas and Moral Theory*. Oxford: Oxford University Press.
- McDowell, J. (1979). Virtue and Reason. *Monist*, 62, 331–50.
- Millgram, E. (2002). Commensurability in Perspective. *Topoi* 21, 217–26. <https://doi.org/10.1023/A:1014825624960>
- Murdoch, I. (1991). *The Sovereignty of Good*. London: Routledge Press.
- Railton, P. (1996). The Diversity of Moral Dilemmas. In H.E. Mason (Ed.), *Moral Dilemmas and Moral Theory* (pp. 140–66). Oxford: Oxford University Press.
- Rawls, J. (1999). *A Theory of Justice* (Revised Edition). Cambridge, Mass.: The Belknap Press.
- Raz, J. (1990). *Practical Reason and Norms*. Princeton: Princeton University Press.
- Raz, J. (2000). The Value of Practice. In *Engaging Reason: On the Theory of Value and Action* (pp. 207–17). Oxford: Oxford University Press.
- Richardson, H. (1990). Specifying Norms as a Way to Resolve Concrete Ethical Problems, *Philosophy and Public Affairs*, 19, 279–310.
- Richardson, H. (1995). Beyond Good and Right: Toward A Constructive Ethical Pragmatism, *Philosophy and Public Affairs*, 24, 108–41.
- Saramago, J. (2003). *The Cave* (Trans. by Casta, M.J.). London: Vintage Books.
- Schapiro, T. (2021). *Feeling Like It: A Theory of Inclination and Will*. Oxford: Oxford University Press.
- Smith, M. (1994). *The Moral Problem*. Malden, Mass.: Blackwell Publishing.
- Smith, M. (2013). On the Nature and Significance of Thick and Thin Ethical Concepts. In S. Kirchin (Ed.), *Thick Concepts* (pp. 97–120). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199672349.003.0007>
- Svarvasdottir, S. (1999). Moral Cognitivism and Motivation. *Philosophical Review* 108, 161–219. <https://doi.org/10.2307/2998300>
- Tessman, L. (2015). *Moral Failure*. Oxford: Oxford University Press.
- Thomson, J.J. (1992). On Some Ways in Which a Thing Can Be Good. *Social Philosophy and Policy*, 9, 96–117. <https://doi.org/10.1017/S0265052500001424>
- Toppinen, T. (2004). Moral Fetishism Revisited. *Proceedings of the Aristotelian Society*, 104, 305–13.

- Väyrynen, P. (2013) *The Lewd, the Rude, and the Nasty: A Study of Thick Concepts in Ethics*. New York: Oxford University Press. doi: <https://doi.org/10.1093/acprof:oso/97891919931451.001.0001>
- Velleman, D. (1996). The Possibility of Practical Reason. *Ethics*, 106, 694–726. <https://doi.org/10.1086/233669>
- Walden, K. (2018). Practical Reason Not as Such. *Journal of Ethics and Social Philosophy*, 13, 124–153. doi: [10.26556/jesp.v13i2.257](https://doi.org/10.26556/jesp.v13i2.257)
- Wallace, J. (2006a). The Rightness of Acts and the Goodness of Lives. In *Normativity and the Will: Selected Essays on Moral Psychology and Practical Reason* (pp. 300–21). Oxford: Oxford University Press.
- Wallace, J. (2006b). Moral Reasons and Moral Fetishes: Rationalists and Anti-rationalists on Moral Motivation. In *Normativity and the Will: Selected Essays on Moral Psychology and Practical Reason* (pp. 322–42). Oxford: Oxford University Press.
- Wiggins, D. (1975-76) Deliberation and Practical Reason. *Proceedings of the Aristotelian Society, New Series* 76: 29–51.
- Wiggins, D. (1998) *Needs, Value, Truth*. New York: Oxford University Press.
- Williams, B. (1965). Ethical Consistency. *Proceedings of the Aristotelian Society, Supplementary Volumes* 39, 103–38.
- Williams, B. & Smart, J.J.C. (1973). *Utilitarianism For & Against*. Cambridge: Cambridge University Press.

How to cite this article: Buss, S. (2022). Personal ideals and the ideal of rational agency. *Philosophy and Phenomenological Research*, 1–23. <https://doi.org/10.1111/phpr.12917>