Immorality and Irrationality

Alex Worsnip
UNC-Chapel Hill

Forthcoming in *Philosophical Perspectives*
Penultimate version; please cite published version where possible

Does immorality necessarily involve irrationality? The question is often taken to be among the deepest in moral philosophy. The project of vindicating an affirmative answer, it is thought, is one of a grandeur and ambition comparable to giving a demonstration that we have free will, or that the mental cannot be reduced to the physical, or that every event must have a cause. There is supposed to be a deep and significant faultline between those who say 'yes' – Kant and Kantians being the primary example – and those who say 'no' – Hume and Humeans being the primary example.

But apparently deep questions sometimes – though not always – admit of deflationary answers. One way of opening up such deflationary answers is to disambiguate one or more of the terms that feature in the question. And the term 'irrationality' is ripe for disambiguation. Specifically, numerous philosophers (including myself) have recently urged a distinction between *structural* rationality and *substantive* rationality.[1] Very roughly speaking, structural rationality is about one's attitudinal mental states standing in relations of coherence to one another. It involves (for example) having consistent beliefs, intending the means to one's ends, having transitive preferences, and respecting one's own beliefs about what one's reasons favor. Substantive rationality, by contrast, is about correctly responding to one's reasons. Call the view that these are two genuine, fundamentally distinct kinds of rationality *dualism* about rationality, and philosophical strategies that appeal to it dualist strategies.

I won't be defending dualism here (at least not directly); I do this elsewhere.[2] Rather, I'll be exploring how it opens up the way for a sensible – I will not say boring – moderate view about the relationship between immorality and irrationality: roughly, that immorality involves substantive irrationality, but not structural irrationality. The purpose of this paper is to defend this moderate view. Interestingly, we'll see that many of the arguments for less moderate views turn either on missing the distinction between substantive and structural rationality, or on misconstruing it.

The plan is as follows. §1 tries to get a slightly more precise fix on our central question, and to clear away some distractions. §2-3 defend the view that (gross) immorality involves substantive irrationality against two different lines of resistance: respectively, the view that morality does not necessarily give us reasons, and the view that though morality does give us reasons, these reasons are not among those that bear on substantive irrationality. §4 defends the view that immorality does *not*

For helpful discussions and comments, I'm grateful to Tom Dougherty, Daniel Fogal, Chris Howard, Nevin Johnson, Geoff Sayre-McCord, and especially Sarah Stroud. I'm also grateful to audiences at the fourth Uppsala Varieties of Normativity Workshop and at the University of Wisconsin-Milwaukee, and to the participants in my Fall 2019 graduate seminar at UNC.

[1] See, among others, Scanlon (2007), Chang (2013), Wallace (2014), Neta (2015), Worsnip (2018b, 2018c), Fogal (forthcoming).

[2] See esp. Fogal & Worsnip (ms.), and Worsnip (ms.: esp. chs. 1 & 5).

necessarily involve structural irrationality against some ambitious neo-Kantian arguments that try to show that it does. §5 concludes.

1. Sharpening the question

In the rough formulation of the question that I began with, I asked whether immorality *necessarily* involves irrationality. The 'necessarily' here reflects the fact that all parties to the debate will concede that immorality *sometimes* involves irrationality. For one thing, it might be that in acting immorally, one (intentionally) does something that one knows will frustrate one's ends, which almost everyone will agree involves irrationality.[3] In some cases, this irrationality is independent of the moral status of the action in question – as when breaking some promise of mine is immoral, and I also know that it will lead to retaliation from the promisee that will prevent me from achieving some end of mine. In these cases, the immoral action (viz., breaking my promise) involves me in irrationality, but in a way such that its status as immoral and its status as irrational are orthogonal. In other cases, however, the two statuses are not so independent. Most obviously, I might have the *de dicto* end of complying with the moral requirements that apply to me, and I might correctly grasp that some moral requirement applies to me, and I will then be irrational if I fail to comply with this moral requirement.

Immoral actions might also be irrational in ways that go beyond instrumental irrationality. For example, if I myself believe that I *all-things-considered* ought not to perform some immoral action, but (intentionally) perform it anyway, I am akratic in a way that – most will concede – is irrational.

All this shows that the view that it is sometimes irrational to be immoral is too weak to be the site of the main controversy here. But conversely, the unqualified claim that it is necessarily irrational to be immoral seems rather strong, in at least two respects.

First, the claim that it is necessarily irrational to be immoral seems to presuppose that moral considerations always override all other sorts of considerations (viz. prudential considerations, aim-given considerations, etc) with respect to what one all-things-considered ought to do. For, if there are some situations in which moral considerations can be outweighed with respect to what one all-things-considered ought to do, then it seems implausible to claim that it would be irrational to act against those moral considerations in such a case. Now, while the claim that moral considerations are always overriding (with respect to what one all-things-considered ought to do) is not an absurd claim by any means, it seems orthogonal to the main question we want to investigate here. It is evidently possible to think that moral considerations always *bear* on verdicts of rationality and irrationality – regardless of, say, whether one cares about acting morally – without signing up to the overridingness claim.[4]

---

[3] Some (e.g. Broome 2013: 151) hold that strictly speaking, *actions* (or absences thereof) can never be irrational; only *intentions* (or absences thereof) can be. But even if this is so, there is surely a derivative sense in which an action can be irrational, or at least "involve" irrationality, when performed intentionally. Similarly, given the right conditions, a failure to act reveals or perhaps even constitutes an absence of an intention to so act (as Broome affirms, *ibid.*), and so can be said to be irrational or involve irrationality.

[4] Although, it is worth noting that at least in many of the most plausible cases in which moral considerations are overridden, it is somewhat unnatural to describe the action that one has most all-things-considered reason to do as 'immoral', as opposed to merely as 'morally sub-optimal' or similar. If this is so, then it might still be OK to say that *immorality* in some stronger sense always involves irrationality, even if there are some cases where moral considerations are overridden.

Second, consider putative cases in which one is *non-culpably* ignorant of (or mistaken about) what morality requires of one: say, cases where one has misleading evidence about what the requirements of morality are. Not everyone agrees that there are such cases.[5] But if there are, it is doubtful that it is irrational to be immoral in such cases.[6] As a general matter, one is not irrational for failing to respond to reasons that one is non-culpably ignorant of.[7] The standard cases used to bring this point out involve descriptive ignorance. To take a common example, if you have no idea that the fish on the plate in front of you contains salmonella, and had no way of knowing this, then you are not irrational for eating the fish, even though the fact that the fish contains salmonella is a decisive reason not to eat it. The lesson of this is that even *substantive* rationality requires responsiveness only to "possessed", "available", or "evidence-relative" reasons. Now, it's at least quite plausible that this also means that it's not irrational to fail to respond to considerations that one may be aware of, but that one cannot recognize *as reasons* because of non-culpable normative ignorance, including non-culpable moral ignorance. Again, while this could be resisted, resolving this debate seems orthogonal to the main question we are investigating here. Allowing for exceptions in cases of non-culpable moral ignorance is consistent with the spirit of the view that moral considerations bear on what is rational or irrational, regardless of (say) whether one cares about them.

Note, however, that it's crucial here that the moral ignorance in question be non-culpable. That is, it must be begotten by misleading evidence or the like, and not by epistemic irrationality or brute insensitivity to what is morally salient. If it's conceded that *just any* moral ignorance lets one off the hook, rationally speaking, then this is a bigger concession, one that collapses the view on offer into the view that we're only irrational for failing to live up to our own moral standards, and never for immorality *as such*.

Given the various points just made, we can state the view that we want to examine more precisely as follows:

(BEARING) Cases of non-culpable ignorance aside, moral considerations necessarily bear on what it is rational to (intend to) do.[8]

If we want to put the view in terms of a connection between *im*morality and *ir*rationality, as we did in our initial imprecise statement of it, then I suggest that we use the term 'gross immorality' to stipulatively restrict ourselves to cases where the moral considerations against some action are strong

[5] Cf. Wedgwood (2019); Smithies (ms.).

[6] Note that this is orthogonal to the (perhaps more controversial) question of whether one might still be *blameworthy* for acting immorally in such cases. This question is discussed by, among others, Harman (2011).

[7] Cf., e.g., Broome (2007: 352-3); Kiesewetter (2017: 160-4); Lord (2018: 8-9).

[8] There may be ways of putting the view that take account of non-culpable moral ignorance without treating these as "exception" cases that merely need to be put to one side. For example, we might introduce the notion of an *evidence-relative* moral demand, where what is morally demanded of one in the evidence-relative sense is relativized not just to one's (potentially misleading) evidence about the relevant descriptive facts, but also to one's (potentially misleading) evidence about the moral facts themselves. (For a persuasive case for the *moral* relevance of moral uncertainty, see Rosenthal forthcoming.) We could then say that *evidence-relative* moral demands, or considerations, necessarily bear on what it is rational to (intend to) do, without the need for setting any cases aside as exceptions.

enough to outweigh any non-moral considerations in favor of performing it, from the standpoint of what one all-things-considered ought to do. Then, we can state the view as follows:

(INVOLVEMENT) Cases of non-culpable ignorance aside, gross immorality necessarily involves irrationality.

How should we (more descriptively) label the view captured by (BEARING) and (INVOLVEMENT)? One candidate is the term 'moral rationalism', which is sometimes used for something in the neighborhood of the view I've been describing.[9] Unfortunately, this term is also used for a bewildering variety of other claims,[10] including at least the following:

Moral Rationalism$_{\text{Reasons}}$. If you are morally required to Φ, then necessarily, you have a reason to Φ.[11]

Moral Rationalism$_{\text{Decisive-Reasons}}$. If you are morally required to Φ, then necessarily, you have a decisive reason to Φ.[12]

Moral Rationalism$_{\text{Epistemology}}$. Moral truths are knowable through reason (as opposed to, e.g., perceptually).[13]

Moral Rationalism$_{\text{Psychology}}$. Moral judgments are psychologically based in faculties of reason (as opposed to, e.g., sentiment or emotion).[14]

Following the subscripting practice above, we might call the refer to the version of moral rationalism that we're interested in – that which is captured by (BEARING) and (INVOLVEMENT) above – as 'Moral Rationalism$_{\text{Rationality}}$'.

In what follows, I will set Moral Rationalism$_{\text{Epistemology}}$ and Moral Rationalism$_{\text{Psychology}}$ aside. While they may have some bearing on Moral Rationalism$_{\text{Rationality}}$, if they do it is an indirect and non-obvious one.[15]

---

[9] E.g., by Doyle (2000), Van Roojen (2010), and Cholbi (2011).
[10] See Schroeter, Jones & Schroeter (2018) for an overview.
[11] E.g., Smith (1994: 62, 2018); Shafer-Landau (2003: 170, ch. 8). We might want to tinker with this formulation – and the next – to rule out views on which the reason that one has to Φ when one is morally required to Φ is not itself a (fundamentally) moral one; e.g., views on which you always have a reason to do what you're morally required to do because, and only because, you will be tortured in hell if you don't, and you have a prudential reason to avoid being tortured to death. Such views arguably don't count as forms of moral rationalism in an interesting sense.
[12] E.g., Portmore (2011: ch. 2).
[13] E.g., Peacocke (2004); Dancy (2006).
[14] E.g., Nichols (2002); Kennett (2006). There is also a debate in early modern philosophy between rationalists and sentimentalists that seems to concern both, or not distinguish between, the psychological and epistemological formulations: for a useful overview, see Gill (2007).
[15] For example, it might be claimed that because morality is *a priori* (as Moral Rationalism$_{\text{Epistemology}}$) claims, everyone ought to be able to grasp its fundamental truths, that because of this, everyone has reasons to comply with it, and that because of *this*, it is irrational not to comply with it. But an argument of this form is controversial at every stage.

By contrast, there is *prima facie* a more intimate tie between the claims about reasons and those about rationality; indeed, some treat these claims as equivalent, on the assumption that rationality is about responding to reasons.[16] Here the central distinction between substantive and structural rationality that I want to rely on comes into view. *Substantive* rationality is about responding to reasons: thus, unless there is some special reason to treat moral reasons differently from other sorts of reasons (I'll come back to this in §3), a vindication of the claim that there are (necessarily) reasons to be moral is ipso facto a vindication of the claim that moral considerations (necessarily) bear on *substantive* rationality. However, one can hold that there are reasons to be moral without thinking that moral considerations bear on *structural* rationality – which is concerned with coherence – or that there's anything structurally irrational about being moral.

As I previewed at the outset, my boringly moderate view is that Moral Rationalism$_{\text{Rationality}}$ is true on a substantive reading of '(ir)rational', but false on a structural reading of '(ir)rational'. Given that substantive rationality is about responding to reasons, the former claim effectively commits me to the (unambiguous) truth of Moral Rationalism$_{\text{Reasons}}$.[17] By contrast, however, I take no position here on Moral Rationalism$_{\text{Decisive-Reasons}}$, which builds in an overridingness claim of the sort I've already said that I want to stay neutral on.[18]

2. Defending Moral Rationalism$_{\text{Reasons}}$

This section and the next are concerned with bolstering the view that Moral Rationalism$_{\text{Rationality}}$ is true on a *substantive* reading of '(ir)rational': that is, cases of non-culpable moral ignorance aside, it is substantively irrational to be grossly immoral. There are two ways to deny this. The first is to deny Moral Rationalism$_{\text{Reasons}}$: that is, to deny that we necessarily have reasons to do as morality requires. The second is to claim, roughly, that though rationality requires responsiveness to some kinds of reasons, it is somehow not irrational to fail to respond to *moral* reasons. This section engages with the first line of resistance, and the next section engages with the second.

I cannot give a comprehensive defense of Moral Rationalism$_{\text{Reasons}}$ in the space available here. What I want to do, instead, is to try to show that some of the most prominent arguments against Moral Rationalism$_{\text{Reasons}}$ themselves turn on an equivocation between substantive rationality and structural rationality.[19] Thus, clearly drawing the distinction between substantive and structural rationality at least bolsters the case for Moral Rationalism$_{\text{Reasons}}$.

---

[16] E.g., Shafer-Landau (2003: ch. 8). Indeed, even calling the claim about reasons 'moral *rationalism*' seems to presuppose such a connection.

[17] Actually, strictly speaking, it commits me only to a precisification of Moral Rationalism$_{\text{Reasons}}$ with an exception-clause for cases of non-culpable moral ignorance, as in the statement of Moral Rationalism$_{\text{Rationality}}$. But in fact, I'm inclined to accept Moral Rationalism$_{\text{Reasons}}$ in its unqualified form, since I think there is at least some good sense of 'reasons' in which one can have reasons that one is non-culpably ignorant of: they just aren't among the reasons that bear on rationality.

[18] For my initial thoughts on overridingness, see Worsnip (2018a: 253-4).

[19] The responses I give in the next two subsections – to Foot and Williams – bear some similarities to the way that Scanlon (1998: 27-29) responds to these same two authors (though he is responding to a different Foot article, and does not use the terminology of substantive and structural rationality here).

(a) Foot

In her classic "Morality as a System of Hypothetical Imperatives," Philippa Foot argues that moral requirements are not "categorical imperatives" in any strong sense of that term; that is, in any sense in which requirements of (e.g.) etiquette are not categorical. Foot concedes that moral requirements and are categorical in a weak sense, namely that they *apply* to us even if we do not desire to comply with them.[20] But, as she points out, requirements of etiquette, as well as other requirements like "club rules," are categorical in this sense too. The injunction to answer third-person invitations in the third person, for example, doesn't *fail to apply* to those who don't care about obeying it.[21] Consequently, Foot says, those who think that moral requirements are categorical in some sense in which requirements of etiquette are not categorical must have something more, something stronger, in mind by 'categorical'. Foot locates what this stronger notion of categoricity might be in terms of the notion of *reasons*. Specifically, moral requirements would be categorical in this stronger sense if the fact that something was required by morality "in itself [gave] us a reason to act,"[22] or (not quite equivalently), if "moral considerations necessarily [gave] reasons for acting to any man".[23] Foot denies that moral requirements (or, indeed, requirements of etiquette) are categorical in this stronger sense. Thus, she denies Moral Rationalism$_{\text{Reasons}}$.

Why? Foot writes:

> "The fact is that the man who rejects morality because he sees no reason to obey its rules can be convicted of villainy but not of inconsistency. Nor will his action necessarily be irrational. Irrational actions are those in which a man in some way defeats his own purposes, doing what is calculated to be disadvantageous or to frustrate his ends. Immorality does not necessarily involve any such thing." (Foot 1972: 310)

After considering various epicycles, Foot then writes: "the conclusion we should draw is that moral judgments have no better claim to be categorical imperatives than do statements about matters of etiquette."[24] Given the setup – on which requirements of etiquette are weakly, but not strongly, categorical, and strong categoricity is glossed in terms of reasons – it is clear that Foot is drawing the conclusion that Moral Rationalism$_{\text{Reasons}}$ is false. Indeed, she immediately adds: "people may indeed follow either morality or etiquette without asking why they should do so, but equally well they may

---

[20] Foot (1972: 307-8). See Joyce (2001: 37) for the "strong" vs "weak" terminology in interpreting Foot's argument.
[21] *Ibid.*: 308-9.
[22] This first quotation comes from a context in which Foot is discussing what it would be for rules of etiquette to be categorical in the stronger sense, but it's clear she intends the formulation to apply to both etiquette and morality.
[23] The first formulation is slightly more committal than the second. The first formulation assumes that what "gives" (or, perhaps better, constitutes) the reason for acting is *the fact that the act in question is morally required*. Conversely, the second formulation is compatible both with this view and with the alternative view (Dancy 2000: 165-7; Zimmerman 2007: Appendix 2) that the reason for acting is given (or constituted) by whatever facts *make it the case* that the fact that the act in question is morally required. I am convinced by Johnson King (2019) that it is actually fine to talk of the fact that something is required as a reason to do it, and that in the final analysis this need not be thought of as competing with talking of the fact that makes it required as a reason to do it; either way of talking is permissible. Still, it's probably better not to bake this into the formulation of moral rationalism, and as such Foot's second, more neutral formulation is better than her first.
[24] *Ibid.*: 312.

not. They may ask for reasons and may reasonably refuse to follow either if reasons are not to be found."[25]

Thus, Foot seems to be arguing from the claim that it isn't necessarily irrational to violate moral requirements, to the claim that they are not (strong) categorical imperatives. The suppressed premise here, obviously, is that if moral requirements *were* (strong) categorical imperatives, it *would* be necessarily irrational to violate them. That is, the structure of the argument is as follows:

(1) If moral requirements were (strong) categorical imperatives (i.e., if Moral Rationalism$_{\text{Reasons}}$ were true), then it would be necessarily irrational to violate them.
(2) It isn't necessarily irrational to violate moral requirements.
So,
(C) Moral requirements are not (strong) categorical imperatives (i.e., Moral Rationalism$_{\text{Reasons}}$ is false).

But this argument equivocates on 'irrational'. Remember what it is for a requirement to be a strong categorical imperative, on Foot's gloss: this involves having *reasons* to comply with it. Thus, if premise (1) is read as using 'irrational' to mean '*structurally* irrational', it would say that if there were reasons to comply with moral requirements, then it would be *structurally* irrational to violate them. But at least *prima facie*, not every failure to respond to one's reasons involves one in the kind of incoherence that amounts to structural irrationality. Thus, 'irrational' as it occurs in premise (1) must be read as referring to *substantive* irrationality.[26]

However, in arguing for premise (2), Foot explicitly appeals to an essentially *structural* notion of irrationality: "irrational actions are those in which a man in some way defeats his own purposes, doing what is calculated to be disadvantageous or to frustrate his ends."[27] And premise (2) is indeed plausible, on a structural reading of 'irrational' (more on this in §4). But as far as I can see, Foot has given us no argument at all for the claim that violating moral requirements does not involve any *substantive* irrationality. Thus, her argument equivocates, relying on a substantive reading of 'irrational' in premise (1), but a structural reading of 'irrational' in premise (2). It thus fails to put any pressure on Moral Rationalism$_{\text{Reasons}}$.

---

[25] *Ibid.*

[26] Even then, premise (1) is questionable as stated, since one might have a reason to comply with a strong categorical imperative that is nevertheless outweighed, and in that case violating the requirement might involve no substantive irrationality either. One option in response to this problem would be to strengthen the definition of 'categorical' still further so that categorical requirements are those that agents necessarily have *overriding* reasons to comply with. But this would reduce the significance of Foot's conclusion: she would only have argued that agents don't necessarily have *overriding* reasons to comply with moral requirements, not that they don't necessarily have reasons to comply with moral requirements. A better option would be for Foot to modify premise (1) to say that if moral requirements were (strong) categorical imperatives, then it would at least *often* be irrational to violate them *irrespective of one's desires*. This is plausible (given a substantive reading of 'irrational'), and with corresponding adjustments to premise (2), gets the argument back on track temporarily.

[27] In fact, if anything, this notion seems even narrower than the notion of structural irrationality properly understood. Foot is picking out means-end incoherence in particular, but there are plausibly other kinds of structural irrationality or incoherence besides means-end incoherence.

(b) Williams

Bernard Williams equivocates in an almost identical way in his "Internal and External Reasons." Williams is arguing for reasons internalism, according to which there cannot be "external" reasons, namely (roughly speaking) reasons that an agent can have to Φ even in the absence of any motive that would be served by her Φ-ing (or any "sound deliberative route" from one's existing motives to a motive that would be served by her Φ-ing).[28] On the assumption that one might not have any motive that would be served by one's doing what is morally required of one, this delivers the falsity of Moral Rationalism$_{\text{Reasons}}$. Williams writes:

> "There are of course many things that a speaker may say to one who is not disposed to Φ when the speaker thinks that he should be, as that he is inconsiderate, or cruel, or selfish, or imprudent; or that things, and he, would be a lot nicer if he were so motivated. Any of these can be sensible to say. But one who makes a great deal out of putting the criticism in the form of an external reason statement seems concerned to say that what is particularly wrong with the agent is that he is *irrational* […] This suggestion, once the basis of an internal reason claim has been clearly laid aside, is bluff." (Williams 1981: 110-1; his italics)

Williams's argument here closely resembles Foot's. He seems to be arguing as follows:

(1) If there were external reasons to do (e.g.) what is morally required, then it would (necessarily) be irrational to fail to do what is morally required.
(2) It isn't (necessarily) irrational to fail to do what is morally required.
So,
(C) There are not external reasons to do what is morally required.

Again, together with the plausible further assumption that there are not always *internal* reasons to do what is morally required – ones that can be explained in terms of its serving one's motives – (C) gives us the result that Moral Rationalism$_{\text{Reasons}}$ is false.

But once again, the argument equivocates. (1) is only plausible given a substantive reading of 'rational': it's not true that a failure to respond to external reasons would have to be *structurally* irrational.[29] But (2) is only intuitively clear given a structural notion of rationality: it's in *that* sense that it's immediately intuitively plausible that the inconsiderate, cruel, selfish person need not be irrational, and that charging them with such irrationality would be "bluff". Williams has not given us an argument that such a person isn't *substantively* irrational. Thus, whichever reading of 'irrational' we adopt, at least one of the premises is seriously dubious (and unargued for). So it is hard to see what pressure Williams's observations here put on the possibility of external reasons.

[28] For the "sound deliberative route" terminology, see Williams 1995: 35.
[29] Cf. also McDowell (1998: 110), though he does not draw the substantive/structural distinction. Indeed, it is plausibly part of what would make a putatively external reason *external*, rather than internal, precisely that failure to respond to it need *not* involve *structural* irrationality.

Admittedly, this is not the only argument against the possibility of external reasons that Williams offers: he has another, much more complex argument.[30] Williams's more complex argument begins with the principle that if A has a reason to Φ, it must be possible (in principle) for that reason to figure in some correct explanation of A's Φ-ing (102). But, Williams points out, an external reason for Φ-ing cannot *by itself* offer an explanation of A's Φ-ing, since by definition an external reason if supposed to be one that is had irrespective of one's motivations (106). Rather, Williams thinks, the most plausible way that such an external reason could (in principle) figure in an explanation of A's Φ-ing is via A's coming to *believe* that she has this external reason (107-9). Moreover, Williams thinks, the external reasons theorist must hold that in coming to believe that she has this external reason, A is "considering the matter aright" (109).[31] And this, Williams thinks, commits the external reasons theorist to the following claim:

(X) A has an external reason to Φ only if it's true that "if the agent rationally deliberated, then, whatever motivations he originally had, he would come to be motivated to Φ". (109)[32]

But, Williams says, once we see this, it becomes implausible that there are any external reasons. The simplest way to read Williams here is as simply claiming that there is no Φ such that, if one rationally deliberated, then *whatever motivations one originally had*, one would come to be motivated to Φ. But a more complex way to read Williams, which I think fits slightly better with his text, has Williams posing a dilemma for the externalist here. Consider what it would take for it to be true that, *were* the agent to rationally deliberate, she would come to be motivated to Φ. One way for that to be true would be if the agent has existing motivations such that a process of rational deliberation (AKA a "sound deliberative route") *from* those motivations would eventuate in a motivation to Φ. But if that is so, then the reason to Φ is in fact an internal one, not an external one. (This remains true even if, somehow, it turns out that there is some Φ such that, for any existing motivations one might have, a process of rational deliberation from those motivations would eventuate in a motivation to Φ. This would just be a universally-shared internal reason.) But how else could it be true that were the agent to rationally deliberate, she would come to be motivated to Φ? It seems that the only other available model is one on which one can rationally deliberate to a motivation to Φ *without taking any existing motivation, as it were, as an input*.[33] But *this*, Williams thinks, is not possible (he calls its impossibility

[30] My reconstruction of this argument is more complicated than the argument that is standardly attributed to Williams, according to which he directly argues that reasons must be able to motivate, that external reasons cannot motivate, and thus that there are no external reasons. Like Finlay (2009), I do not find the argument in the text to be this simple; I have tried to stay closer to the text than this common reconstruction does, giving page references for each step of the argument. My reconstruction also differs from Finlay's own interpretation, but settling all the interpretative issues here definitively would take me too far off track.

[31] Why? Because the reason is only genuinely external if it was already there before the agent came to believe that she had it; and if that's so, then coming to believe this is "considering the matter aright".

[32] Williams doesn't say how this is supposed to be handle *pro tanto* reasons that are outweighed or defeated. Maybe the idea is that motivations are also *pro tanto*, and one should always have *some* motivation to Φ in the presence of *some* reason to Φ. This doesn't seem obviously right to me, but I set it aside.

[33] Given this reading, the argument requires a distinction between deliberation that proceeds 'from' a motivation in a robust sense where that motivation plays a genuine explanatory role as an "input", and deliberation that proceeds perhaps *in the presence of* some motivation, but does not proceed *from* that motivation, on the other. The sharpness of this distinction might be questioned, but I won't press that point here.

"Hume's basic point").[34] Thus, Williams offers the externalist a dilemma: *either* the purported external reason collapses into an internal reason, *or* the ascription of the external reason rests on a faulty model of rational deliberation.

This argument also, I think, equivocates on the notion of rationality, as it appears in (X) (in the phrase 'rational deliberation'). Consider first a purely structural notion of rationality. Given a structural notion of rationality, rational deliberation only involves eliminating any structural irrationality one might have in one's existing attitudes, where this includes forming the attitudes that one's existing attitudes "commit" one to given the norms of structural rationality but which one does not yet have. For example, if one intends some end and believes some means is necessary for the end, structurally rational deliberation would involve coming to intend the means (or giving up the end, or the means-end belief). Given this gloss on rational deliberation, however, the external reasons theorist should deny that she is committed to (X). The external reasons theorist need not agree that whenever one has a reason to Φ, one's existing attitudes already commit one, by norms of structural rationality, to a motivation to Φ, or that correction of one's existing attitudes for merely *structural* irrationality would necessarily eventuate in a motivation to Φ. This is compatible with recognition of Williams's claim that the external reasons theorist must hold that, when one comes to recognize that one has some external reason to Φ, one is "considering the matter aright". The claim is just that one can come to recognize legitimate normative demands on oneself that do not amount to *commitments* generated by structural rationality from one's existing attitudes.

By contrast, suppose that we employ a *substantive* notion of rationality. On this notion of rationality, rational deliberation will involve more than just drawing out the commitments of one's existing attitudes: it will involve one's being appropriately sensitive to the reasons that one has. Then it is more plausible that the external reasons theorist is committed to something like (X).[35] However, given a substantive gloss on rationality, the external reasons theorist should deny what Williams calls "Hume's basic point", namely that rational deliberation can only involve deliberating *from* one motivation to another. For, on the substantive conception, rational deliberation also involves being appropriately responsive to one's reasons. And when one's reasons support Φ-ing, this involves recognizing that fact and forming a motivation to Φ in response to it.[36] Given the connection between reasons and substantive rationality, to say that *substantive* rationality cannot require one to Φ in the absence of some existing motivation that would be served by one's Φ-ing *just is* to deny that one can have external reasons, and is thus blatantly question-begging.[37]

---

[34] This might be explained in terms of the claim that it is psychologically impossible to arrive at a motivation through a process of deliberation that does not begin with some other motivation. Or it might be explained in terms of the claim that no such deliberative process could count as *rational* (still less be rationally *required*, as it would need to be in order for it to be determinately true that were the agent to deliberate rationally, she would arrive at the motivation to Φ), because motivations (and transitions to them) cannot be rationally evaluated, except in terms of one's other, pre-existing motivations: they cannot be rationally required *ex nihilo*. I am not sure which of these two explanations Williams intends.

[35] *Modulo* the concerns about pro tanto reasons mentioned in fn. 32 above.

[36] Maybe this transition doesn't deserve to be called *deliberation*. But then the external reasons theorist can just insist that (X) should be reformulated to refer to the motivations that one would have were one to be rational, rather than were one to rationally *deliberate* specifically. Cf. McDowell (1998: 99-100) for discussion.

[37] What if the claim is that it is simply *psychologically impossible* to form a motivation to Φ in response to a recognition of one's reasons to Φ, absent some antecedent motivation that would be served by Φ-ing? This too seems like something that the external reasons theorist should simply reject. At best (for Williams), there is simply a standoff here.

(c) The irrationality of morality?

I'll highlight one final line of argument again Moral Rationalism$_{\text{Reasons}}$ that likewise trades on ignoring the substantive/structural rationality distinction.[38] Suppose that Villanelle intends to kill Eve Pelastri in the most violent way possible, and believes that the most violent way to kill Eve is to kill her using a cleaver.[39] Intuitively, there is something wrong with Villanelle if she does not intend to kill Eve using a cleaver. Obviously, this wrongness is not moral in character. Rather, it seems natural to say that, given the ends she has, Villanelle would be irrational if she did not intend to kill Eve using a cleaver.[40] But, assuming that rationality is a matter of responding to one's reasons, this suggests that Villanelle must have decisive reasons to (intend to) kill Eve using a cleaver.[41] But obviously, this is a grossly immoral act. Thus, it seems, it must be that one can have decisive reasons to intend to do something grossly immoral, when one intends something to which that act is a necessary means.

This does not quite show that Moral Rationalism$_{\text{Reasons}}$ is false, since it could be that Villanelle has moral reasons not to (intend to) kill Eve using a cleaver that are outweighed. But a generalization of the case will suggest that whenever one has some end such that the believed means to that end are immoral, any moral reasons against intending or taking the means to that end will be outweighed. And that suggests that moral reasons are weak enough that they are always outweighed by instrumental concerns to the contrary. At this point, it's not clear what the point of affirming that there are such moral reasons would be; it seems cleaner to just deny Moral Rationalism$_{\text{Reasons}}$.

However, once again, distinguishing substantive and structural rationality blocks this argument. With this distinction in view, we can agree that, given her existing end and means-end belief, there *is* a good sense in which Villanelle would be irrational if she failed to intend to kill Eve with a cleaver. However, the relevant irrationality here is structural: the problem with Villanelle would be that she exhibits a kind of incoherence. Since structural irrationality is not about responding to reasons, this entails nothing to the effect that Villanelle has decisive reasons to (intend to) kill Eve with a cleaver, merely because she intends to kill Eve in the most violent possible (and believes that the most violent way possible to kill Eve is with a cleaver). Nor does it entail that she lacks decisive moral reasons to refrain from all this murderous behavior (or, indeed, that it would not be *substantively* irrational for her to engage in it).[42] This, the argument against Moral Rationalism$_{\text{Reasons}}$ is blocked.

---

[38] It's hard to find a definitive statement of this line of argument, but it seems to have been "in the air" for a long time. Darwall (1983: 43 n. 1) attributes something like it to Harman (1976), but I don't find a clear statement of it in the latter. My response to this line of argument has something in common with Darwall's: see in particular Darwall (1983: 14-16, ch. 4; 2016: 260-3).

[39] The case is adapted and updated from Darwall (1983: 15).

[40] The example as stated involves strict instrumental (ir)rationality, which deals only with resolved intentions and full beliefs to the effect that some means are *necessary* for some end. But it can also be run using norms of subjective utility maximization, which proceeds in light of one's credences and preferences. So, Villanelle might have credences and preferences such that it would be irrational for her not to (intend to) kill Eve with a cleaver. I think that norms that relate such credences and preferences to choice are also best viewed as (candidate) structural requirements.

[41] Assuming that (right-kind) reasons to intend to Φ are always also reasons to Φ (cf., e.g., Hieronymi 2005), it also suggests that Villanelle has similar reasons not just to intend to kill Eve with a cleaver, but actually to do so.

[42] Does this diagnosis mean that there is a *conflict* between structural and substantive rationality here, with structural rationality requiring Villanelle to intend to kill Eve with a cleaver, and substantive rationality requiring her not to intend to kill Eve with a cleaver? Not necessarily, if the relevant norm of structural rationality is "wide-scope", merely

(d) Dualism explains the intuitions better

The pattern over the last three subsections has been that once we make the dualist move of distinguishing substantive and structural rationality, we can block arguments against Moral Rationalism$_{\text{Reasons}}$. However, at the same time, we can do justice to some of what motivated these arguments. An opponent of Moral Rationalism$_{\text{Reasons}}$ who did *not* distinguish substantive and structural rationality would be unable to account for the intuition that there is a sense of 'irrational' in which immorality need not amount to irrationality, as Foot and Williams stressed. And similarly, she would be unable to account for the intuition that there's a sense of 'irrational' in which (given her background states) Villanelle would be irrational if she did *not* intend to kill Eve with a cleaver. This irrationality goes beyond the substantive unreasonableness of her initial end of killing Eve in the most violent way possible: it is a further mistake, and one of a fundamentally different kind.

So far, however, I've only really shown that the dualist can provide an *alternative* explanation of these intuitions that doesn't require us to reject Moral Rationalism$_{\text{Reasons}}$. Now, I want to suggest that the dualist's explanation of the intuitive data is actually *superior* to that of the denier of Moral Rationalism$_{\text{Reasons}}$. As we've seen, one major piece of intuitive data is that there are a range of cases in which it's natural to say that there's at least a good sense of 'irrational' in which immorality does not amount to irrationality. But, I'll suggest, this cannot always be explained by denying Moral Rationalism$_{\text{Reasons}}$.[43]

What are the cases in which it is natural to deny that immorality amounts to irrationality? One prominent case is that of the amoralist, who doesn't care about being moral. Broadly internalist accounts of reasons like those endorsed by Foot and Williams – which go along naturally with the rejection of Moral Rationalism$_{\text{Reasons}}$ – have an explanation of why such a person doesn't have any reason to be moral, and thus (assuming a connection between reasons and rationality) of why it isn't irrational for them to disobey morality's commands. But consider a different sort of case, involving an agent who does care about being moral, but is simply mistaken about morality's demands. For instance, consider:

> **Conscientious Meat-Eater.** Ben has a strong (*de dicto*) desire to avoid doing anything morally impermissible. Suppose for the sake of argument that it is morally impermissible to eat meat, because meat-eating contributes significantly to the suffering of animals. Ben is aware that meat-eating contributes significantly to the suffering of animals. However, Ben honestly (though falsely) believes that the suffering of animals is not morally weighty, and so nevertheless believes that it is morally permissible to eat meat. Consequently, Ben goes ahead and (intentionally) eats meat.

forbidding the incoherent combination of intending an end, believing a means is necessary for that end, and not intending the means, without saying which particular one of these three incoherent states ought to be revised. Cf., among many others, Broome (1999, 2013: ch. 8).

[43] Here I am reprising, in a somewhat different form, a point I made in my (2016).

I submit that intuitively, there's at least a sense of 'irrational' in which Ben is not irrational for (intentionally) eating meat. Moreover, it seems to me that this intuition is no weaker than the intuition that the amoralist is not irrational. To whatever extent that it's not irrational to ignore moral demands when you don't care about morality generally, it's also not irrational to fail to comply with some specific moral demand that you don't consider to be a genuine one.

However, notice that Ben has a desire to avoid doing what is morally impermissible. That should be enough to give him a reason not to eat meat, even by the lights of those who subscribe to reasons internalism or some related doctrine. Now, admittedly, as I noted in §1 above, one is not (even substantively) irrational for failing to respond to reasons that one is *non-culpably* ignorant of. But suppose that Ben's ignorance of his moral duty not to eat meat is culpable: he is unjustified in believing that meat-eating is permissible; he has no specially misleading evidence to the effect that is permissible. Still, it seems to me that there is a sense in which, given that Ben believes it is permissible to eat meat, he is not irrational in (intentionally) eating meat. But denying Moral Rationalism$_{\text{Reasons}}$, and subscribing to the kind of reasons internalism that tends to go along with that denial, gives us no explanation of how this is so.

By contrast, the dualist move of isolating a distinctive notion of structural rationality does give us such an explanation. Notwithstanding the reasons to eat meat that Ben has, and his culpability in his ignorance of these reasons, Ben is still (at least as described so far) structurally rational: he has a coherent – if substantively mistaken or unjustified – moral framework, which he follows consistently. But note that once we accept dualism, we no longer need to deny Moral Rationalism$_{\text{Reasons}}$ in order to explain why the amoralist is in one good sense rational, either. Thus, the intuition that it often isn't irrational to be immoral doesn't end up lending any support to the denial of Moral Rationalism$_{\text{Reasons}}$.

Let's take stock of §2 as a whole. Nothing I've said in this section proves that Moral Rationalism$_{\text{Reasons}}$ is true. But I do think that Moral Rationalism$_{\text{Reasons}}$ should be the default position in the debate, in the absence of compelling arguments against it. Remember that the debate about Moral Rationalism$_{\text{Reasons}}$ is about whether, when it's a fact that you are morally required to Φ, you necessarily have a reason to Φ. The debate about Moral Rationalism$_{\text{Reasons}}$ thus presupposes that there are at least some facts about what's morally required of us. If this is common ground, then I think the burden of proof is on the denier of Moral Rationalism$_{\text{Reasons}}$ to explain what is special about the normative concept of a reason that means that facts capable of generating moral requirements, and corresponding 'ought's, don't also generate or constitute *reasons* for action.[44] This is exactly what Williams (for example) is attempting to do in giving an argument for the constraint that to count as a reason, some fact must be appropriately related to our motives. If the available arguments for these kinds of constraints fail, then we should presumptively accept Moral Rationalism$_{\text{Reasons}}$.

[44] It won't do, for example, to insist that there's something spooky or anti-naturalistic about reasons that float free of our desires or ends, while just presupposing that there's nothing similarly spooky or anti-naturalistic about moral *requirements* or 'ought's that float free of our desires or ends. The relevant difference between reasons on one hand, and requirements or 'ought's on the other, needs to be explained. Moreover, we need an explanation of what is special about *morality* that means that it generates requirements without necessarily generating reasons, where for other normative domains such as epistemic normativity these seem to come together.

Here I have been trying to show that some of the most prominent such arguments do fail, specifically because they equivocate between substantive and structural rationality. But there are of course other arguments, both possible and actual, and a full defense of Moral Rationalism$_{\text{Reasons}}$ would need to engage with them.[45] I will satisfy myself with the conclusion that, *ceteris paribus*, drawing the distinction between substantive and structural rationality strengthens the position of defenders of Moral Rationalism$_{\text{Reasons}}$ in the debate about whether it is true.

3. Moral reasons bear on substantive (ir)rationality

Recall that I am currently defending the view that Moral Rationalism$_{\text{Rationality}}$ is true on a substantive reading of 'rational'. One line of resistance to this view, considered in the last section, denies Moral Rationalism$_{\text{Reasons}}$. A different lines of resistance concedes Moral Rationalism$_{\text{Reasons}}$, but nevertheless holds, roughly speaking, that it need not be substantively irrational to fail to respond correctly to one's moral reasons. This is only a rough way of putting the strategy, because (as I defined it in §1) Moral Rationalism$_{\text{Rationality}}$ already concedes that there are some cases in which one can be immoral without substantive irrationality. Specifically, it concedes this both in cases of non-culpable moral ignorance (if there be such cases) and in cases where moral reasons are outweighed (if there be such cases). So this line of resistance needs to go beyond noting these particular exceptions.

Prima facie, it is hard to see what could justify this strategy. When it comes to other kinds of reasons – such as prudential reasons, or (even more clearly) epistemic reasons, we take it as axiomatic, as a platitude, that failures to respond to such reasons amount to substantive irrationality.[46] Why should moral reasons be any different? The tendency to take it as axiomatic that substantive rationality requires responsiveness to these other sorts of reasons, but not to moral reasons, appears to be little more than anti-morality bias.[47]

But perhaps, when we get into the details of exactly how to understand substantive (ir)rationality in terms of reasons-responsiveness, it will fall out of our best theory that it is often not irrational to fail to respond to one's moral reasons. I'll examine two reasons-responsiveness accounts of substantive rationality that have this consequence. But I'll contend that the moves they make that excuse immoral agents from (substantive) irrationality simultaneously exclude many other intuitively irrational agents from (substantive) irrationality, making their overall theories of substantive (ir)rationality too forgiving.

(a) Lord

[45] For example, that of Manne (2014), which does not equivocate on the notion of rationality. For a reply to Manne, see Brunero (2017).

[46] For the epistemic case, compare e.g., Williamson (2000: 164); Kelly (2006: §2); Cohen (2013: 99).

[47] My rejection of an asymmetry between moral and non-moral reasons on this count goes back to my earlier (2016). However, at that time, I myself didn't accept dualism, and was inclined to a purely structural theory of rationality simpliciter. So I argued that we should deny that rationality requires responsiveness to either moral or even prudential or epistemic reasons. Though my view on this has changed, the underlying conviction that there's no genuine asymmetry between the moral and non-moral cases has not.

The first account is that given by Errol Lord (2018). Lord holds that (substantive) rationality requires you to respond only to the reasons that you "possess". The reasons for Φ-ing that you *possess* are, for Lord, only a subset of the reasons that there *are* for you to Φ. Possessing a reason, according to Lord, involves meeting two conditions. The first is an epistemic condition: roughly, it says that, where R is some reason to Φ, you need to be in a position to know R. It seems hard to dispute that reasons that fail this condition (or, at least, something quite close to it[48]) do not bear on verdicts of (ir)rationality: as we noted earlier, you are not irrational for failing to respond to facts that you are not even in a position to know. However, our formulation of Moral Rationalism$_{\text{Rationality}}$ already accounts for this by setting cases of non-culpable ignorance aside. Thus, Lord's first condition does not put any pressure on Moral Rationalism$_{\text{Rationality}}$ as I've formulated it.

However, Lord thinks that the first condition on possessing a reason is not enough to rule out all the reasons that shouldn't bear on verdicts of (ir)rationality. He also introduces a second, "practical" condition on possessed (and thereby, rationality-relevant) reasons, which states that some reason R to Φ is one that you possess only if you're in a position to manifest know-how about how to use it as a reason. Having this know-how, Lord maintains, requires you to be at least somewhat disposed to Φ when the relevant consideration is a reason to Φ. Since amoralists and those who are otherwise deeply morally ignorant (including those who are *culpably* morally ignorant) won't be disposed to act as the moral reasons recommend, they won't count as having the relevant know-how, and thus won't be rationally required to respond to these moral reasons, nor irrational when they fail to do so. Lord embraces this consequence.[49]

This problem with adding this second condition, however, is that it makes Lord's theory too underdemanding in a range of other cases where verdicts of (substantive) irrationality are extremely natural: indeed, cases that seem like paradigm examples of substantive irrationality.[50] Consider, say, a flat-earth conspiracy theorist. Intuitively – assuming that the conspiracy theorist lives in a society broadly like ours, and is aware of the unanimous scientific consensus that the world is not flat – his evidential reasons decisively support believing that the world is not flat, and he is substantively irrational for failing to believe this. However, it's easy to imagine such a person so that he fails to meet Lord's practical condition for "possessing" these reasons. A true-believing flat-earth conspiracy theorist is not disposed to believe that the earth is not flat in the presence of the relevant evidential reasons for so believing – for example, in the presence of testimony from scientific experts that the earth is not flat. For he believes that these scientific experts are part of the conspiracy, and as such is not disposed to believe what they tell him. (Indeed, he might even be actively disposed to believe the *opposite* of what they tell him: their testimony might only make him *more* confident that the earth is

[48] There's room for disputing around the margins here. Perhaps, for example, there are considerations that you're not in a position to know, but whereby your not being in a *position* to know them is nevertheless culpable, and you're still irrational for failing to respond to these reasons.

[49] See Lord (2019), where he makes this consequence of his view explicit, and then writes, echoing the passage from Williams quoted in §2b above, "there are still many other ways in which we can condemn the amoralist. We can look at him funny, speak harshly to him, lock him up, and banish him from our social circles."

[50] I have also developed this objection, slightly differently and in a somewhat different context, in Fogal & Worsnip (ms.: §6.1).

flat.)[51] But if this is so, the true-believing conspiracy theorist is not in a position to manifest know-how about how to use this testimony as a reason. Thus, he does not "possess" it as a reason in Lord's senses (and similarly for all the other evidential reasons for believing that the earth is not flat). Thus, on Lord's account, he is not irrational for failing to respond to it: his failure to believe that the earth is not flat is not irrational. Similarly for other paradigm cases of (seemingly) irrational doxastic agents, such as climate change deniers, or someone who believes that there are fairies at the bottom of her garden.

This is a bad result. Admittedly, since that the flat-earther has a story about why the scientific experts' testimony is not good evidence – namely, that they are part of the conspiracy – he avoids *incoherence*. But this ought to save him only from the charge of structural irrationality, and not from the charge of substantive irrationality. This substantive sense of 'irrational' is what epistemologists typically have in mind when they say that it is irrational to believe that the earth is flat, or that climate change is a hoax. They don't withdraw this charge if it's clarified that the doxastic agent has some complex, crazy, but internally coherent story about why their evidence really does support believing that the earth is flat or that climate change is a hoax.[52] The agent's beliefs are substantively irrational as long as they are wrong about this – that is, as long as their evidence in fact *doesn't* support believing that the earth is flat, or that climate change is a hoax (no matter whether they *believe* that it does). I think the epistemologists are clearly right in thinking that there's at least *a sense* in which such beliefs, and indeed the absence of beliefs to the contrary, are irrational, and we should want our theory of *substantive* rationality to capture it.[53]

Moreover, at a more general and theoretical level, Lord's introduction of the practical condition seems to restrict substantive irrationality to those who are *disposed* to do as their reasons support doing, but fail to manifest this disposition. This seems oddly narrow. If anything, those who are not even *disposed* to do as their reasons support doing seem to be even worse off, rationally speaking.

Therefore, I think that we should reject Lord's practical condition.[54] It thus fails to tell against the view that it is substantively irrational to fail to respond to one's moral reasons.

(b) Kiesewetter

---

[51] It might be replied that he only needs to have *some*, possibly weak, possibly defeated, disposition to believe that the earth is not flat in response to the evidence, and that such weak dispositions come cheaply so that any realistic flat-earther will have them. But then the same can be said about those who fail to respond to their moral reasons, and so Lord's account will no longer save them from charges of irrationality either.

[52] Some extreme subjective Bayesians are an exception here. But at the risk of being tendentious, these extreme subjective Bayesians defy common sense. According to this view, as long as I have a high prior that pink elephants will invade China conditional on its raining tomorrow, it's a fully rational response to my seeing it rain tomorrow to come to have a high credence that pink elephants will invade China. This is incredible. In the common sense picture, there are at least some facts about what a given body of evidence supports believing, and it's (substantively) irrational to believe against what one's body of evidence supports believing, no matter what one's priors are.

[53] Strikingly, Lord claims to be vindicating the standard epistemological picture of rationality over the purely coherentist conception that is more popular in the literature on practical rationality (Lord 2018: 4-5). But if I'm right, his introduction of a practical condition for reason-possession into his theory of rationality means that it's subject to similar objections to the purely coherentist conception.

[54] Lord has an independent argument for introducing his extra constraint on rationality-relevant reasons. See Fogal & Worsnip (ms.: §6.3) for a response.

The second account I'll consider is given by Benjamin Kiesewetter (2017). Kiesewetter holds that (substantive) rationality requires one to correctly respond to one's "evidence-relative" reasons, where there is no Lord-style practical condition on what it is for a reason to be evidence-relative. However, perhaps surprisingly, Kiesewetter also holds that a failure to correctly respond to one's evidence-relative reasons – that is, to do what (substantive) rationality requires of one – does not suffice for *ir*rationality. Rather, he claims that "irrationality occurs in the particular case where this failure is *guaranteed* by the agent's attitudes" (*ibid.*: 236). As I understand it, what this means is that some set of attitudes *S* (where *S* can be a singleton set) is irrational iff there is no possible world where every attitude in *S* is supported by one's (evidence-relative) reasons. Here we are not holding what one's evidence-relative reasons are fixed across possible worlds. Effectively, the claim is that for some set of attitudes *S* to be irrational, there has to be no possible totality of evidence-relative reasons that one could have such that each attitude in *S* would be supported by those reasons.

This is rarely true of immoral intentions to act. For example, there's presumably *some* possible world in which one's evidence-relative reasons *do* support (intentionally) eating meat – say, some world where meat-eating does not cause any suffering, or where doing so is somehow necessary to prevent some even greater suffering. Thus, Kiesewetter's account does not count the intention to eat meat (or the intentional action of eating it) as irrational – not even at the actual world, where one's evidence-relative reasons decisively support not intending to eat meat. For the intention to eat meat to be irrational at the actual world, it would have to be *guaranteed* to constitute a failure to respond to one's evidence-relative reasons, where that means that there is no possible world in which it is supported by one's evidence-relative reasons. Again, Kiesewetter embraces this consequence.[55]

Like Lord's account, Kiesewetter's account withholds verdicts of irrationality in other cases, where they are very natural. Presumably there is some possible world where one's evidence-relative reasons support believing that the earth is flat, and so by Kiesewetter's account, such a belief is not irrational – not even at the actual world, where one's evidence-relative reasons decisively tell against this belief. Again, the same holds for climate change deniers, fairy-believers, and indeed almost any paradigmatically (substantively) irrational belief. Again, these are bad results: these are paradigm cases of substantive irrationality, and any theory that doesn't label them as such seems too underdemanding and forgiving. Moreover, at a more general and theoretical level, Kiesewetter's view entails that if an attitude (or set of attitudes) is irrational in *one* possible world, it is irrational in *all* possible worlds. It's thus deeply unsuited to capturing the vast majority of instances of substantive irrationality, where an attitude is irrational at the actual world because of how one's reasons and evidence are at the actual world, and not because there's no possible world in which the attitude could be supported.

Which attitudes *do* come out as irrational on Kiesewetter's view? At least according to Kiesewetter, the combinations of attitudes usually associated with *structural* irrationality come out as irrational on his view.[56] For example, Kiesewetter thinks, there is no possible world in which one's evidence-relative reasons support *both* (say) the belief that it's raining *and* the belief that it's not raining. Thus, this *combination* of attitudes comes out as irrational on his view. But Kiesewetter also thinks that

[55] Kiesewetter 2017: 236.

[56] See Worsnip (ms.: ch. 5) for some doubts about whether this claim is true in full generality.

certain individual attitudes come out as irrational on his view. He gives Parfit's case of a person who is indifferent to pain on future Tuesdays as an example (*ibid.*: 238). Kiesewetter thinks there is no possible world in which this attitude is supported by one's (evidence-relative) reasons, and so it counts as irrational on his account.

This raises a further question, namely whether there are some cases of immoral intentions that *do* come out as irrational on Kiesewetter's view, contrary to his own ambitions. And it seems to me that there might well be. For example, plausibly there is no possible world in which one's evidence-relative reasons support intending to torture babies for fun.[57, 58] Now, to be clear on the dialectic, I don't myself think it's a bad result to say that this intention is irrational, but to the extent that Kiesewetter himself wanted to say that immoral intentions are not irrational, it is a problem for him.

Moreover, independently of one's intuitions about whether immoral intentions should count as irrational, Kiesewetter's view has a strange feature. Which immoral intentions are irrational, for Kiesewetter, will track not *how severe* the immorality – or failure of reasons-responsiveness – is, but rather how *modally robust* it is. To see how these might come apart (this is just an illustration), consider a view that combines agent-neutral consequentialism with the view that moral reasons are overriding.[59] Suppose further that in the actual world, starting a nuclear war would have devastatingly terrible consequences (and that your evidence indicates this). Intending to do so is thus a severe failure of reasons-responsiveness. Nevertheless, there's some possible world in which starting a nuclear war has net beneficial consequences, and so where intending to do so is not a failure of reasons-responsiveness. So, the intention to start a nuclear war, on Kiesewetter's view, is not irrational (even at the actual world). By contrast, consider the action of choosing one's relative's slightly lesser happiness over some stranger's slightly greater happiness. Intending to make such a choice is only a very slight failure of reasons-responsiveness. But – if agent-neutral consequentialism is (necessarily) true, and moral reasons are (necessarily) overriding – it is a failure of reasons-responsiveness in every possible world – and so the intention is, on Kiesewetter's view, irrational. This *combination* of verdicts – that (intentionally) starting the nuclear war involves no irrationality but that the (intentional) minor favoring of one's relative over a stranger does involve irrationality – seems to me perverse, independently of disputes about the plausibility of the two verdicts taken individually.

---

[57] At least if we're restricting ourselves to right-kind reasons for intention, as Kiesewetter is.

[58] There is a difficult issue here, analogous to the problem about the specificity of maxims often raised for Kant's moral theory, about the level of detail at which an intention, or the act that is the object of that intention, should be specified here. One and the very same intention could be described as an intention *to torture a baby*, or as the intention *to torture a baby for fun*, or as the intention *to torture a baby in a circumstance where doing so produces no good consequences*. And this can make a difference for the verdict Kiesewetter's theory yields: whereas there are at least arguably some possible worlds in which one's evidence-relative reasons support intending to torture a baby (say, where doing so is somehow necessary to prevent the destruction of the universe), there are no possible worlds in which one's evidence-relative reasons support intending to torture a baby *for fun*, or intending to torture a baby *in a circumstance where doing so produces no good consequences*. This vagueness might allow Kiesewetter to play around with the descriptions of each intention or action to get the results he wants in each case, but in the absence of some general criterion for the level of specificity at which intentions or actions are to be described, this looks unprincipled. Indeed, the indeterminacy of the results Kiesewetter's account yields, given the unclarity in how specific our descriptions of intentions are to be, can itself be seen as a weakness in his view.

[59] What follows is adapted from a comment I made on a blog post in a discussion of Kiesewetter's book. See http://peasoup.us/2018/07/ndpr-forum-the-normativity-of-rationality/

Taking all of these problems together, I think we should reject Kiesewetter's view that one is only irrational when one has attitudes that (in his sense) *guarantee* a failure of reasons-responsiveness. Of course, Kiesewetter is right that 'irrational' may be a somewhat harsh term for very slight failures of reasons-responsiveness. But there are other, better ways to accommodate this point. We can say that (substantive) irrationality fundamentally comes in degrees, or that in consists in a failure of reasons-responsiveness above some threshold. So Kiesewetter too fails to mount a convincing challenge to the view that it is substantively irrational to fail to respond to one's moral reasons.

(c) Some more general points

I've argued that the qualifications that Lord and Kiesewetter put on their reasons-responsiveness accounts of (ir)rationality should be rejected, primarily on the grounds that they fail to deliver verdicts of (substantive) irrationality in paradigm cases like that of the flat-earther, and are thus too forgiving. With these qualifications rejected, there's nothing to rule out moral reasons from bearing on verdicts of substantive (ir)rationality. But I also think that cases like that of the flat-earther, and the intuitiveness of verdicts of irrationality in such cases, build a more positive case for the view that moral reasons should bear on these verdicts. The case is one by analogy.

The flat-earther is aware of a fact – that the scientific experts say that the earth is not flat – that is a decisive reason to believe the earth is not flat, but fails to recognize it *as a decisive reason to believe the earth is not flat*, and as such isn't disposed to respond to it as such. This failure of recognition helps to save the flat-earther from a charge of structural irrationality, or incoherence, but it doesn't let him off the hook for substantive irrationality.

Now consider someone like the meat-eater we considered in §2d above. This person, we can now note, is analogous to the flat-earther. She is aware of a fact – that meat-eating contributes significantly to the suffering of animals – that is a decisive reason not to eat meat (again, alter the example if you disagree) – but fails to recognize it *as a decisive reason not to eat meat*, and as such doesn't respond to it as such.[60] By parallel, while this failure of recognition helps the save the meat-eater from a charge of structural irrationality, or incoherence, it shouldn't let her off the hook for substantive irrationality – any more than the precisely analogous features of the flat-earther get him off the hook.

Strikingly, neither Lord nor Kiesewetter ultimately wants to distinguish substantive and structural (ir)rationality as two *bona fide* kinds of irrationality, different in kind. I hypothesize that this results in their trying to account for intuitions about both substantive and structural irrationality at once.[61] Lord and Kiesewetter see that there is an intuitively clear sense in which characters like the meat-eater – and, perhaps to a lesser degree, characters like the flat-earther – are not being irrational.[62]

---

[60] And, note, this is true not just of a *conscientious* meat-eater of the kind we considered in §2d above, but also of an amoralist. The amoralist may recognize that the fact that meat-eating contributes significantly to the suffering of animals makes it *morally wrong* to eat meat, but she presumably denies that it is a decisive reason not to eat meat from the all-things-considered point of view.

[61] This is especially evident in Kiesewetter's embracing a reasons-responsiveness theory of rationality (which I associate with *substantive* rationality), but then embracing a theory of *ir*rationality the primary purpose of which is to capture the idea that the patterns of attitudes associated with *structural* irrationality are irrational.

[62] Cf. Kiesewetter's (2017: 236) claim that "criticism of someone as irrational is different from moral *or epistemic* criticism" (my italics).

But since they don't distinguish substantive and structural rationality, they give up on the thought that there's some *other* sense in which such characters *are* being irrational. Especially in the case of the flat-earther, this undermines the whole motivation for wanting a notion of rationality as reasons-responsiveness, one that goes beyond mere coherence, in the first place. By contrast, if we are willing to separate substantive and structural rationality, we can note the *sense* in which the meat-eater and even the flat-earther are not irrational – namely the structural sense – leaving us free to get relatively demanding with our theory of substantive rationality, and to say that both of these characters *are* being substantively irrational.[63] So this is another place where failure, or refusal, to make the distinction between structural and substantive rationality has distorted the debate.

4. Does immorality involve *structural* irrationality?

In the last paragraph, I assumed that the meat-eater is not *structurally* irrational. But this leads me to the other side of my moderate position on Moral Rationalism$_{\text{Rationality}}$ – namely that it is *false* on a *structural* reading of 'rationality'. Some philosophers – most prominently, a group of philosophers working in the Kantian ethical tradition – challenge this. They hold that immorality must involve some kind of incoherence, or structural irrationality.[64] On a plausible reading, this is one of the primary aspects of Christine Korsgaard's version of neo-Kantianism.[65] But a more explicit articulation of the view can be found in Julia Markovits's (2014) book *Moral Reason*.[66] Due to space constraints, this is the version of the view I'll engage here, though I think that the objections I raise generalize to Korsgaard's version of the view as well.

Markovits argues for the following claim:

(HUMANITY) If an agent fails to have humanity as an end, that agent is structurally irrational.

---

[63] This does leave us with a sociological-psychological question about why our intuitions seem to get drawn to *structural* rationality when we reflect on the moral case, but to *substantive* rationality when we reflect on the epistemic case. I tried some initial speculations on this in my (2016).

[64] Was this Kant's own view? Markovits (2014: ch. 4) argues that it was. Indisputably, Kant thought that immorality involves irrationality, but I am not enough of a Kant scholar to assess Markovits's arguments to the effect that he thought that it involves structural irrationality in particular, or how to map the structural/substantive distinction onto the way Kant thinks about rationality.

[65] See e.g. Korsgaard 1996. Chang (2013) reads Korsgaard in broadly this way.

[66] Some clarifications about Markovits's view. First, Markovits's ultimate aim in her book is actually to show that there are (universally shared) *reasons* to be moral – that is, to defend Moral Rationalism$_{\text{Reasons}}$ – and to show that this view can be squared with a form of reasons internalism, which (as we saw in §2 above) is typically taken to entail the falsity of Moral Rationalism$_{\text{Reasons}}$. However, as I explain in the main text below, her strategy for doing this goes *via* the claim that immorality involves structural irrationality. Here, I am not concerned with the validity of this transition from the claim that immorality involves structural irrationality to the claim that we have reasons to be moral, but only with the truth of the former claim.

Second, Markovits uses the term 'procedural rationality' rather than 'structural rationality', but it is clear she intends to pick out what I mean by the latter. The term 'procedural' suggests a *process*-oriented notion of (structural) rationality (cf. Kolodny 2007), but it is not clear that Markovits really intends this, since she often talks about the requirements of "procedural rationality" as if they are synchronic requirements on states (cf. Markovits 2014: 51-2, 110, 113-4, 117). I prefer the term 'structural rationality' since it is more clearly neutral on the state/process dispute.

Having humanity as an end, in the relevant sense, does not involve aiming at maximizing humanity; rather, it involves respecting and valuing the humanity in others. This end is "served" by actions that treat others as end in themselves – that is, by actions that satisfy the second formulation of Kant's categorical imperative. And it is frustrated by actions that treat others as a mere means – that is, by actions that violate Kant's second categorical imperative.

Markovits doesn't completely spell out how (HUMANITY) is supposed to establish the result that all immoral action is structurally irrational. However, I take it that the argument is supposed to go roughly as follows. Markovits thinks, as Kant did, that each formulation of the categorical imperative captures the whole of morality. Thus, she holds:

(COMPLETENESS) All immoral action frustrates the end of humanity.

Finally, the following may seem plausible, and seems to be assumed by Markovits at times:

(INSTRUMENTAL) If an agent (intentionally) performs an action that frustrate her ends, that agent is structurally irrational.

If (HUMANITY), (COMPLETENESS), and (INSTRUMENTAL) are all true, then it follows that all immoral actions (that are performed intentionally) are structurally irrational. To see this, suppose that some agent performs some immoral action (intentionally). By (COMPLETENESS), this action frustrates the end of humanity. Now, either the agent has humanity as an end, or she doesn't. If she does have humanity as an end, then by (INSTRUMENTAL), she is structurally irrational. If she doesn't have humanity as an end, then by (HUMANITY), she is structurally irrational. So either way, she is structurally irrational.

Now, one might question any of the three premises that generate this result. I will be focusing on (HUMANITY), and Markovits's argument for it, because I think that the way that the argument for (HUMANITY) fails is the most philosophically significant, and the most likely to generalize to other attempts to argue that all immoral action is structurally irrational. However, I will briefly comment on (COMPLETENESS) and (INSTRUMENTAL) in turn.

(COMPLETENESS) is of course controversial, to put it mildly. However, if it fails, it fails as a matter of substantive, first-order moral theory: there is no straightforward, clear mistake lying behind it. Moreover, even if it fails, it's possible that there is some other supreme moral end such that any immoral action frustrates that end. And that leaves it open for some argument of the *form* of the one that we're considering here to succeed. I will therefore set aside objections to (COMPLETENESS).

(INSTRUMENTAL) may initially seem like the most uncontroversial of the three premises – it is the one least particular to ambitious Kantian theory. However, there is a significant potential problem with it as it stands. On most views of structural (ir)rationality, you are not *structurally* irrational – not incoherent – for (intentionally) performing some action that *actually* frustrates some end of yours, if you don't yourself *believe* that it frustrates that end of yours.[67] If that is so, (INSTRUMENTAL) is false as stated, and in a way that matters for the argument. Consider someone who has humanity as

[67] Cf., e.g., the formulation of the instrumental requirement in Broome (2013: 159).

an end, and who falsely believes that Φ-ing does not frustrate the end of humanity (e.g., because she falsely believes that Φ-ing does not constitute treating someone as a mere means, when in fact it does). If such a person were to intentionally Φ, she would be acting immorally, but (given the correction to (INSTRUMENTAL)) she would not be structurally irrational.

Perhaps there is some way of patching the argument in response to this problem. The Kantian might hold, for instance, that someone who claims to value humanity, but is seriously mistaken about what it is to respect someone's humanity, or to treat others as ends as opposed to means, doesn't really count as "valuing humanity" in the relevant sense. However, even if there is no such patch, the argument would still establish something that, while weaker than the conclusion that *all* immoral action involves irrationality, is still quite strong and very interesting: namely, that in any instance where one *knowingly* treats someone else as a means, one must be structurally irrational. Indeed, if – as the Kantian likely thinks – to fail to value humanity is already to have an immoral attitude, (HUMANITY) on its own shows that there is at least a *kind* of immorality that ensures structural irrationality. The position I'm inclined to hold denies even these weaker claims. For this reason, I'm going to set the problem with (INSTRUMENTAL) aside as well, and focus on (HUMANITY).

As Markovits recognizes, (HUMANITY) clearly can't just be taken as a *prima facie* plausible starting point: it needs to be argued for. Now, Markovits is clear that on her view – and I agree – structural rationality does not *directly* require agents to have any particular positive attitude, including moral ends such as the end of humanity. Rather, she says, the requirement to have such an end is generated *indirectly* through its relation to one's other ends.[68] As I read her, the idea here is as follows:

(COMBINATION) For any end E that one might have, structural rationality prohibits {having E, lacking the end of humanity}.

Together with the fairly weak and plausible premise that it's constitutive of agency that one have *some* ends – one cannot be an agent while lacking any ends whatsoever – (COMBINATION) entails (HUMANITY). But why accept (COMBINATION)? This is where the real action is in Markovits's argument.

The argument goes as follows. Just as it's incoherent to fail to value the means to an end that you value, Markovits thinks, it's also incoherent to value a means without valuing "the more fundamental end to which it is instrumental" (131). Markovits uses this general principle to argue for a kind of chain of principles that ultimately show it to be structurally irrational to have any end while lacking the end of humanity. To illustrate this, Markovits takes an arbitrary example of an end: the end of flossing on a regular basis.[69] Markovits then argues as follows:

- It's structurally irrational to {value flossing on a regular basis, not value good dental health}
- It's structurally irrational to {value good dental health, not value pain prevention (or a longer life)}

[68] See Markovits (2014: 51-2, 110).

[69] Markovits treats having an end of X as interchangeable with valuing X: so having the end of flossing on a regular basis is, for her, the same as *valuing* flossing on a regular basis.

- It's structurally irrational to {value pain prevention (or a longer life), not value myself}
- It's structurally irrational to {value myself, not value humanity}

Each of these claims is supposed to be an application of the general principle that it's structurally irrational, or incoherent, to value a means without valuing the more fundamental end to which it is instrumental. But anyone who values flossing on a regular basis and doesn't value humanity must be structurally irrational in at least one of the above ways. Thus, such a person must be structurally irrational. And the same sort of argument is supposed to generalize for any end one might have. Whatever one's ends are, they must ultimately bottom out in one's valuing humanity.

I think this argument fails, and not just in the details. The overarching problem, as I'll explain shortly, is that none of the above combinations of states have the right features to be structurally irrational. While there may be *some* kind of problem with these combinations of states, if there is, it isn't one of structural irrationality.

This may seem like a surprising claim to make. Sometimes one hears the notion of a structural requirement of rationality glossed as follows: some requirement is a requirement of *structural* rationality just if it prohibits *combinations* of attitudes, as opposed to individual ones.[70] But this is not adequate. There are prohibitions on combinations of attitudes that are not requirements of structural rationality.[71]

Suppose that you're trying to decide what to do tonight. You've been thinking about proposing marriage to your partner some time soon. Let's suppose that your reasons (and hence, substantive rationality) permit, but don't require, you to intend to propose tonight. Let's suppose also that your reasons (and hence, substantive rationality) permit you to intend to watch a gory horror movie with your partner tonight. Suppose that either plan would be a good one. Still, it might be a terrible idea to *both* watch a gory horror movie with your partner *and* propose marriage to your partner on the same night. In that case, your reasons, and hence substantive rationality, prohibit {intending to propose to your partner tonight, intending to watch a gory horror movie with your partner tonight}. Still, you would not be structurally irrational, or incoherent, for having both intentions. (Maybe you, falsely and substantively irrationally, think that doing both would be awesome.) The two intentions, we might say, are *combinatorically unreasonable*, rather than structurally irrational.

Closer to home for the debate we're considering, there are some – most notably Gensler (1985) – who hold that there are some *moral* prohibitions on combinations of attitudes. On one reading of the Golden Rule, for example, it forbids {intending to do A to another person, desiring that she doesn't do A to you in like circumstances}. It's clearly not conceptually confused to think that this is a moral requirement, but nevertheless to deny that it is structurally irrational. Thus, again, it can't follow straightaway from some requirement's being a prohibition on a combination of attitudes that it is a requirement of structural rationality.

This makes it clear that even if there's something wrong with (say) the combination {valuing flossing on a regular basis, not valuing good dental health}, the problem need not be that the two

[70] Cf., e.g., Cohen (2013: 109).

[71] Arguably, there are also some requirements of structural rationality that prohibit individual attitudes. Broome (2013: 153) gives the example of the requirement not to believe a conjunctive proposition of the form (*p* & not-*p*).

states are jointly structurally irrational. But why positively think that the problem is *not* one of structural irrationality? Because, I suggest, we should accept the following conceptual constraint on a theory of structural rationality:

> (CONSTRAINT) Judgments about the structural (ir)rationality of a combination of states should not depend on – that is, should be independent of – substantive judgments about whether there are good reasons for holding those states, or what would constitute a good reason for holding them.[72]

To see the appeal of (CONSTRAINT), consider some paradigm instances of structural irrationality. One can see that it is structurally irrational to {believe that all taxes are illegitimate, believe that sales taxes are legitimate} without having to make any judgments about which, if either, of these two beliefs is in fact substantively supported by good reasons, or what would constitute a good reason for holding either belief. That's why you and I could agree that this combination is structurally irrational even if we totally disagree on which (if either) of these beliefs is substantively reasonable, and why one can see that the combination is structurally irrational even if one is totally agnostic on the reasonableness of the two beliefs. Similarly, one can see that it is structurally irrational to {intend not to break the strike, believe that grading papers constitutes breaking the strike, intend to grade papers} independently of any judgment about which of these states are substantively reasonable: again, people of wildly different views about the reasonableness of each of the three states could agree on this, as could someone with no views on their reasonableness.

But now consider the combination {valuing flossing, not valuing good dental health}. This combination, I suggest, looks bad only because, and to the extent that, we recognize that the only plausibly good *reason* to floss on a regular basis, or to value doing so, is that it promotes one's dental health. We'll recognize this only if we find it implausible that there is good reason to value flossing on a regular basis intrinsically, and think that there are no other plausible values that regular flossing instrumentally promotes.[73] Thus, the judgment that there is something bad about this combination is *not* independent of substantive judgments about the (would-be) reasons for the states involved. Thus, by (CONSTRAINT), the combination is not structurally irrational. (And similarly, I think, for the combinations of attitudes at *every* step of Markovits's chain argument.[74]) Thus, Markovits's argument for (HUMANITY) fails.[75]

---

[72] Cf. also Kiesewetter (2017: 236).

[73] Markovits comes close to conceding this when she qualifies her claim that it's irrational to value flossing without valuing good dental health by saying that this holds "in the absence of other reasons for regularly flossing" (131). So whether it is irrational to value flossing without valuing good dental health depends on whether there are other reasons for regularly flossing, which is a judgment about substantive reasons. Similarly, she writes that "if I value flossing, I must value good dental health, because it's the only (plausible) source of the value of flossing" (144). But this judgment of what is *plausible* as a source of value is a judgment about substantive reasons.

[74] Though it need only be true at one step in the chain for the argument to fail.

[75] The problem remains if Markovits says only that the combinations of attitudes she highlights are *imperfectly* structurally rational rather than irrational. This is sometimes how she talks (see, e.g., 132). It makes no difference, since it is still letting one's judgments about structural rationality be partly determined by one's judgments about reasons.

Notably, this isn't a problem only for the particular combinations of states that Markovits identifies. Rather, it is a problem for the general claim, which her argument trades on, that it's structurally irrational, or incoherent, to value a means without valuing the more fundamental end to which it is instrumental. For the question "to what more fundamental end is this means instrumental?" is a question that belongs to the domain of substantive judgments about reasons. In effect, it asks what end can provide a good *reason* to pursue or value the means in question. For example, the claim that good dental health is the "more fundamental end" to which regular flossing is instrumental – which is assumed in order to generate the supposed structural irrationality of valuing the latter without valuing the former – is effectively just the claim that the need to maintain good dental health provides good reason to engage in regular flossing, and that other considerations do not. Even if this claim is true, one can imagine a perfectly structurally rational agent who disputes it, and values regular flossing as a means to some other end, or even for its own sake. Such an agent would be mistaken about her reasons, not structurally irrational.

Now, what plausibly *would* be structurally irrational is {valuing regular flossing *as a means to* good dental health, not valuing good dental health}. More generally, if talk of "the more fundamental end" were to be interpreted as picking out the end that is more fundamental *according to the agent herself* – that is, the end for the sake of which *she* values the relevant means, then it might be true that it's structurally irrational to value some means without valuing the "more fundamental end". But on this interpretation of such a principle, there's no reason to think that it will ultimately require the agent to value humanity in particular, rather than just requiring her to value *whatever* it is that she herself values other things for the sake of. Markovits thinks that the end of humanity is special, because it's the end that can provide the most fundamental justification for other ends, and which doesn't itself need justification by some further end. But that, again, is a substantive judgment about reasons.[76] As such, again by (CONSTRAINT), it can't justify the verdict that an agent who fails to value humanity fails to be structurally rational.

Obviously, the objection I've been making here turns on (CONSTRAINT). Someone might worry that it is too restrictive, or that, since "structural (ir)rationality" and "coherence" are terms of art, Markovits is free to employ a broader notion of structural (ir)rationality to which (CONSTRAINT) doesn't apply. In fact, however, I'll now argue that given Markovits's theoretical ambitions, she herself must accept (CONSTRAINT).

Markovits ultimately wants to *use* the claim that structural rationality requires us to hold moral ends to argue that we all have reasons to comply with moral demands, even on a broadly internalist view of reasons. The kind of internalist view Markovits favors says, roughly, that one has reason to Φ iff Φ-ing serves some end that one *would* have after one's existing ends are corrected for structural irrationality. Thus, if structural rationality itself requires us to have the moral required ends – like, as (HUMANITY) claims, the end of humanity – then we have reasons to do whatever is necessary to achieve those ends, and hence, to act morally.

[76] As becomes even clearer in her arguments for this claim (see esp. 135-6).

The crucial point here is that structural rationality is supposed to be the *more* fundamental notion – both metaphysically and epistemologically – in terms of which the notion of a reason is understood and demystified.[77] Markovits is quite explicit about this. She writes:

> "The notion of rationality that plays a central role in the internalist account of reasons is *procedural* [i.e. structural[78]], not *substantive* […] The standards of procedural rationality may be both less controversial and […] significantly harder to question than the substantive standards of rationality to which externalists appeal […] The procedural standard of rationality, if not exactly uncontroversial, may nonetheless be one that someone who disagrees with the internalist at the outset about what her *reasons* are might agree on. So it could serve as a kind of Archimedean point against which we might brace ourselves in disputes about reasons. Externalists, by contrast, must appeal to a substantive standard – one that simply incorporates, as a *rational requirement*, the need to respond to the very reason whose existence their interlocutor disputes." (55; her italics)

If judgments about what is required by structural rationality themselves turn on contentious disputes about what is really a reason for what, or what can plausibly explain the value of what, then the role that Markovits wants such judgments to play as an "Archimedean point" in disputes about reasons is clearly compromised. We would no longer be able to use them in order to resolve disagreements about reasons, and nor would be able to claim to be explaining the notion of a reason in turns on the notion of structural rationality. Indeed, we would be doing exactly what Markovits accuses externalists of in the final sentence of the above quotation. For these reasons, it seems that Markovits is herself committed to (CONSTRAINT).

Moreover, this isn't an idiosyncratic fact peculiar to Markovits. It's precisely the sort of ambition Markovits has that makes the neo-Kantian project of showing that it's structurally irrational to be immoral so alluring. This manifests in Markovits's promise to be able to square that claim that morality gives us reasons with a kind of reasons internalism. But it also manifests in the promise to provide us with a new moral epistemology that reveals that moral requirements we're under by starting with what ends we need to have on pain of incoherence, rather than with a faculty of moral intuition. And it manifests in the claim to be able to vindicate the normativity of moral requirements in terms of the claim that complying with them amounts to structural irrationality or incoherence. All of these ambitions are compromised if we use the notion of structural irrationality simply to smuggle in substantive claims about reasons.

Thus, on any notion of structural rationality *not* subject to (CONSTRAINT), it's not clear what's really achieved by showing that it's "structurally irrational" to be immoral. But, as I've argued, once we accept (CONSTRAINT), the attempt to show this fails. And that is the fundamental problem with the neo-Kantian position of which Markovits is an exemplar.

5. Conclusion

[77] This can be seen as an attempt to reduce substantive rationality to structural irrationality. For discussion of this aspect of Markovits's view, see Worsnip (ms.: ch. 4).

[78] See fn. 66 above.

In this paper, I've defended the moderate view that it is substantively irrational, but not structurally irrational, to be (grossly) immoral. Thus, drawing the substantive/structural distinction opens up the way for a broadly deflationary answer to the debate about whether it is irrational, in some undifferentiated sense, to be immoral. Not only this, but numerous existing contributions to that debate rest on ignoring or equivocating on this distinction (Foot, Williams), refusing to fully acknowledge it (Lord, Kiesewetter), or neglecting a key feature of it – namely, (CONSTRANT) (Markovits). The broader methodological upshot is that the distinction between substantive and structural rationality is not just a piece of taxonomical housekeeping. Rather, it allows us to make progress in substantive debates, such as that about moral rationalism, and to make them more tractable.

References

Broome, J. (1999). "Normative Requirements," *Ratio*, 12: 398-419.

-------- (2007). "Does Rationality Consist in Responding Correctly to Reasons?," *Journal of Moral Philosophy*, 4/3: 349-374.

-------- (2013). *Rationality Through Reasoning*. Chichester: Wiley-Blackwell.

Brunero, J. (2017). "Recent Work on Internal and External Reasons," *American Philosophical Quarterly*, 54/2: 99-118.

Chang, R. (2013). "Grounding Practical Normativity: Going Hybrid," *Philosophical Studies*, 164/1: 163-187.

Cholbi, M. (2011). "The Moral Conversion of Rational Egoists," *Social Theory and Practice*, 37/4: 533-556.

Cohen, S. (2013). "A Defense of the (Almost) Equal Weight View," in Christensen & Lackey (eds.), *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press.

Dancy, J. (2000). "Should We Pass The Buck?," *Royal Institute of Philosophy Supplement*, 47: 159-173.

-------- (2006). "On How to Be a Moral Rationalist," *Philosophical Books*, 47/2: 103-110.

Darwall, S. (1983). *Impartial Reason*. Ithaca: Cornell University Press.

-------- (2016). "Making the "Hard" Problem of Moral Normativity Easier," in Lord & Maguire (eds.), *Weighing Reasons*. Oxford: Oxford University Press.

Doyle, J. (2000). "Moral Rationalism and Moral Commitment," *Philosophy and Phenomenological Research*, 60/1: 1-22.

Finlay, S. (2009). "The Obscurity of Internal Reasons," *Philosophers' Imprint*, 9/7.

Fogal, D. (forthcoming). "Rational Requirements and the Primacy of Pressure," *Mind*.

Fogal, D. & Worsnip, A. (ms.) "Which Reasons? Which Rationality?"

Foot, P. (1972). "Morality as a System of Hypothetical Imperatives," *Philosophical Review*, 81/3: 305-316.

Gensler, H.J. (1985). "Ethical Consistency Principles," *Philosophical Quarterly*, 35/139: 156-170.

Gill, M. (2007). "Moral Rationalism vs. Moral Sentimentalism: Is Morality More Like Math or Beauty?," *Philosophy Compass*, 2/1: 16-30.

Harman, E. (2011). "Does Moral Ignorance Exculpate?," *Ratio*, 24: 443-468.

Harman, G. (1976). "Practical Reasoning," *Review of Metaphysics*, 29/3: 431-463.
Hieronymi, P. (2005). "The Wrong Kind of Reason," *Journal of Philosophy*, 102/9: 437-457.
Johnson King, Z. (2019). "We Can Have Our Buck and Pass it Too," *Oxford Studies in Metaethics*, 14: 167-188.
Joyce, R. (2001). *The Myth of Morality*. Cambridge, UK: Cambridge University Press.
Kelly, T. (2006). "Evidence," *Stanford Encyclopedia of Philosophy*.
Kennett, J. (2006). "Do Psychopaths Really Threaten Moral Rationalism?," *Philosophical Explorations*, 9/1: 69-82.
Kiesewetter, B. (2017). *The Normativity of Rationality*. Oxford: Oxford University Press.
Kolodny, N. (2007). "State or Process Requirements?," *Mind*, 116/462: 371-385.
Korsgaard, C.M. (1996). *The Sources of Normativity*. Cambridge, UK: Cambridge University Press.
Lord, E. (2018). *The Importance of Being Rational*. Oxford: Oxford University Press.
-------- (2019). Reply to Nathan Howard's *Ethics* Review of *The Importance of Being Rational*. Online at PEA Soup; available at http://peasoup.us/2019/07/ethics-review-forum-lords-the-importance-of-being-rational-reviewed-by-howard/.
Manne, K. (2014). "Internalism about Reasons: Sad But True?," *Philosophical Studies*, 167/1: 89-117.
Markovits, J. (2014). *Moral Reason*. Oxford: Oxford University Press.
McDowell, J. (1998). *Mind, Value & Reality*. Cambridge, MA: Harvard University Press.
Neta, R. (2015). "Coherence and Deontology," *Philosophical Perspectives*, 29: 284-304.
Nichols, S. (2002). "How Psychopaths Threaten Moral Rationalism," *The Monist*, 85/2: 285-303.
Peacocke, C. (2004). "Moral Rationalism," *Journal of Philosophy*, 101/10: 499-526.
Portmore, D. (2011). *Commonsense Consequentialism: Wherein Morality Meets Rationality*. Oxford: Oxford University Press.
Rosenthal, C. (forthcoming). "What Decision Theory Can't Tell Us About Moral Uncertainty," *Philosophical Studies*.
Scanlon, T. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
-------- (2007). "Structural Irrationality," in Brennan, Goodin, Jackson & Smith (eds.), *Common Minds: Themes from the Philosophy of Philip Pettit*. Oxford: Oxford University Press.
Schroeter, F., Jones, K. & Schroeter, L. "Introduction," in Jones & Schroeter (eds.), *The Many Moral Rationalisms*. Oxford: Oxford University Press.
Shafer-Landau, R. (2003). *Moral Realism: A Defence*. Oxford: Oxford University Press.
Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell.
-------- (2018). "Three Kinds of Moral Rationalism," in Jones & Schroeter (eds.), *The Many Moral Rationalisms*. Oxford: Oxford University Press.
Smithies, D. (ms.). "The Problem of Morally Repugnant Beliefs."
Van Roojen, M. (2010). "Moral Rationalism and Rational Amoralism," *Ethics*, 120/3: 495-525.
Wallace, R.J. (2014). "Practical Reason," *Stanford Encyclopedia of Philosophy*. Available at https://plato.stanford.edu/entries/practical-reason/
Wedgwood, R. (2019). "Moral Disagreement and Inexcusable Irrationality," *American Philosophical Quarterly*, 56/1: 97-108.

Williams, B. (1981). "Internal and External Reasons," in his *Moral Luck*. Cambridge, UK: Cambridge University Press.
-------- (1995). "Internal Reasons and the Obscurity of Blame," in his *Making Sense of Humanity*. Cambridge, UK: Cambridge University Press.
Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.
Worsnip, A. (2016). "Moral Reasons, Epistemic Reasons, and Rationality," *Philosophical Quarterly*, 66/263: 341-361.
-------- (2018a). "Eliminating Prudential Reasons," *Oxford Studies in Normative Ethics*, 8: 236-257.
-------- (2018b). "Reasons, Rationality, Reasoning: How Much Pulling-Apart?," *Problema*, 12: 59-93.
-------- (2018c). Review of Benjamin Kiesewetter's *The Normativity of Rationality*, *Notre Dame Philosophical Reviews*.
-------- (ms.) *Fitting Things Together: Coherence and the Demands of Structural Rationality*.
Zimmerman, M.J. (2007). "The Good and the Right," *Utilitas*, 19/3: 326-353.