Utilitas (2025), 1–17
doi:10.1017/S0953820825100101

CAMBRIDGE UNIVERSITY PRESS

ARTICLES

Structural Rationality in Desire

Alex Gregory

The University of Southampton, Southampton, UK
Email: a.m.gregory@soton.ac.uk

(Received 28 February 2024; revised 12 June 2025; accepted 20 June 2025)

Abstract

Can desires be irrational? This paper focuses on the possibility that desires might be irrational because they fail to cohere with other mental states of the person in question. Recent literature on structural irrationality has largely neglected structural requirements on desire, and this paper begins to rectify that neglect. This paper endorses various rational requirements on desire, but primarily focuses on the instrumental requirement to desire the means to our ends. It explains how this requirement should be understood, and defends it from numerous objections, such as the worry that there are no real instrumental desires but only combinations of ultimate desires and beliefs, and the worry that it would require us to desire even very foolish means to our ends.

Keywords: structural rationality; desire; instrumental rationality; instrumental desire

When might a desire be irrational? One kind of answer to this question might focus on ways in which a desire can fail to track the objective value of its object – e.g. perhaps it is irrational to prefer pain to pleasure. But in this paper, I focus on a different kind of answer, one that focuses on ways in which a desire might be in tension with your other attitudes, so that your irrationality consists in a kind of internal incoherence. On some views, this is the only kind of irrationality there is, and on other nearby views this is at least the only kind of irrationality there is *with respect to desire* (see e.g. Scanlon 1998, 25 for the first view, and perhaps Hume 1738 for the second). But even for those of us who are sympathetic to the existence of multiple kinds of irrationality, internal incoherence is *one* kind of irrationality, and one worth understanding.

The paper is structured as follows: Section 1 introduces the kind of "structural" irrationality that is our focus, and Section 2 highlights one plausible structural requirement on desire: the instrumental requirement to desire the means to your ends. Sections 3-6 defend the instrumental requirement from various objections, thereby vindicating this canonical rational requirement on desire. Section 7 quickly points out two further points of detail. Section 8 returns to the broader topic and posits four further rational requirements on desire, in addition to the instrumental requirement. As this summary indicates, much of the paper de facto focuses narrowly on the instrumental

requirement on desire. This makes sense: it is the most obvious place to begin in investigating rational requirements on desire since it is the canonical such requirement.

Before we begin, a brief note on what I mean by "desire." On some uses of the word, "desires" are always strong influences, with great emotional significance, perhaps ordinarily sexual or otherwise bodily in nature. But that is not the relevant sense here: we are focused on a more general sense of "desire" that includes even modest desires to cross the road, to reply to an email, or turn up the heating (for related discussion, see e.g. Gregory 2021, 6–7; Heathwood 2019; Lauer and Condoravdi 2014, 589–90; Lewis 1988, 323; Nagel 1970, 29; Schueler 1995). Perhaps this use of "desire" is something of a philosophers' term of art, and perhaps the word "want" is closer in meaning to our target. But we'll stick with philosophical tradition and label all these states "desires." In the other direction, some might use "desire" (or more often, "preference"), to mean what the agent *most* desires *on balance*. It should be clear that this use presupposes some prior notion of *pro tanto* desires or preferences that are being combined, and those earlier pro tanto states are our subject matter here. Perhaps there are also rational requirements on overall desire – see section 8 – but we'll treat our primary topic as pro tanto desire.

1 Structural rationality

The kind of irrationality that I have in mind – internal incoherence – is sometimes referred to as *structural* irrationality (e.g. Scanlon 2007; Worsnip 2021). This contrasts with *substantive* irrationality. The distinction is easy to see with other attitudes such as belief: it is structurally irrational to have inconsistent beliefs, and this is different in kind from the substantive irrationality involved in having beliefs that fail to respect the evidence, or the truth. Similarly, with respect to our intentions, we might distinguish the structural irrationality involved in having inconsistent intentions from the substantive irrationality involved in intending something objectively foolish.

Can we say any more to help identify structural irrationality? Alex Worsnip offers three hallmarks of structural rationality (Worsnip 2021, 7–8). First, making a judgement about structural rationality requires no knowledge of the agent's circumstances or evidence: knowledge of their mental states alone suffices to justify the relevant judgements. This contrasts with judgements of substantive rationality, which require us to know something about the agent's circumstances or at least their evidence. Second, judgements about structural rationality can be shared by people who nonetheless have deep substantive disagreements. This is part of the reason why charges of structural irrationality can often be dialectically effective, since they can be made even in the absence of wider agreement on relevant ethical or epistemic truths. Third, judgements about structural rationality hinge on patterns of attitudes considered in relative abstraction from their exact contents, so that it is irrational to believe P and believe ¬P regardless of what we fill in for P. These same three hallmarks will be shared by the requirements on desire that I discuss below.

With very few exceptions, the contemporary literature on structural rationality focuses on the (ir)rationality of beliefs and intentions, saying almost nothing about desire. Why is that? One possible partial explanation is that the canonical kind of structural irrationality is direct inconsistency – as when you believe P and believe ¬P, or intend to φ and intend to not φ – and this kind of inconsistency seems perfectly rational in the case of desire (pace Blackburn 1998, 509–10; Fullhart and Martinez 2024, 324–25). For example, if you desire to go to the dentist, but also desire not to go, that seems perfectly rational: it seems right to feel desideratively conflicted about such a mixed

https://doi.org/10.1017/S0953820825100101 Published online by Cambridge University Press

blessing. And similar remarks apply about even ultimate desire: if you can save either one stranger or another, but not both, it seems right to (pro tanto) desire each of these conflicting outcomes.[1] But although it is true that direct inconsistency can be rational in desire, this leaves open the possibility that other combinations of desire might be irrational. After all, with respect to intention, it is not *just* direct inconsistency that is structurally irrational, but also, say, failing to intend the means to your ends (e.g. Brunero 2020; Broome 2013, 159). I say there is a parallel kind of structural irrationality in desire if you fail to desire the means to your ends. And there are other structural requirements on desire too: broader requirements to render them coherent with our beliefs and intentions. In short, although the permissibility of direct inconsistency in desire might partially explain why our topic has been neglected, it does not justify a continued neglect.

The neglect of desire in discussions of structural irrationality is regrettable; the topic has great importance. For example, much of decision theory might be understood as bearing on questions about structural rationality in desire: we might understand putative constraints on preference orderings – e.g. regarding completeness, transitivity, continuity, and/or independence – as putative requirements of structural rationality (for these notions and further references see Steele and Stefánsson 2020, secs 1–2). It is surprising that the philosophical literature on rationality has largely not been integrated with this existing body of knowledge, and there may be low-hanging fruit for advances in both areas by bringing them together. This paper does not directly address these questions, but it will lay some groundwork for doing so. Or for another example, some Kantians effectively claim that moral requirements can be derived from requirements of structural rationality (e.g. Markovits 2014; Smith 1994). Whether that is possible will depend in part on the breadth of those structural requirements: the more wide-ranging they are, the easier the Kantian task might be. So evaluating the full range of structural requirements of rationality is essential to evaluating the Kantian project.

Two further possible applications of our topic deserve special mention. The first is the debate between reasons internalists, who claim that all reasons for action depend on our desires, and reasons externalists, who claim that at least some reasons for action are independent of our desires (for the former, see Williams 1981; 1995; for the latter, see e.g. Parfit 2011, 58–110). The debate between these camps is sometimes fierce, but we might wish to strike a more conciliatory position. With the distinction between substantive and structural rationality in mind, such a middle-ground view seems possible: we might accept reasons externalism as a truth about substantive rationality, but also accept reasons internalism as part of the truth about structural rationality (Gregory 2021, 192–96; Worsnip 2021, 281–88; see also Scanlon 1998, 366). Whether this conciliatory view is defensible depends on various issues, but one is whether reasons internalism can be plausibly recast as a truth about structural rationality. Whether that is so depends on the nature of structural rationality with respect to desire.

The other notable application of our topic is the wider debate about the very existence of requirements of structural rationality. Some authors have been sceptical about the existence of such requirements, arguing that they are just some kind of shadow cast by

[1]One route to this conclusion goes via the claim that our desires ought to track objective values, or reasons for action. If that is right, then any pluralist view about those things would generate pressure on us to have conflicting desires (see e.g. Gregory 2021, 22; see also Williams 1976; De Sousa 1974). Unless we accept some very stark conflict between substantive and structural rationality in such cases, we are forced to admit that there is no structural requirement to have consistent desires.

https://doi.org/10.1017/S0953820825100101 Published online by Cambridge University Press

requirements of substantive rationality, and nothing more (e.g. Kiesewetter 2017; Lord 2018; Kolodny 2005). I find this view implausible, and prefer to say that there are just two autonomous domains of normative assessment here (for a comprehensive defence, see Worsnip 2021). But if we are uncertain on that topic, the paper helps contribute to it by presenting some more *prima facie* structural requirements that sceptics will need to explain away.[2]

Two final notes: First, perhaps "irrational" sometimes carries special critical force that would be too strong as condemnation for some modest violations of rationality. But this is not how I use "irrational" – when I use "structurally irrational" I just mean "less than perfectly structurally rational," so that such charges need not imply that the relevant failure was particularly egregious. Second: to save words, hereafter whenever I use "irrational" or a cognate term, I always mean structural irrationality. This is just a decision to save on words, not the more controversial view that there is no other kind of irrationality with respect to desire (for helpful related discussion, see Fogal and Worsnip 2021).

2 The instrumental requirement on desire

Everyday talk is rife with the idea of wanting things as means to ends; our theorising about rationality had better make sense of that. As a result, one obvious candidate structural requirement on desire is a requirement to desire the means to your ends. Here is a natural formulation of the Instrumental Requirement on Desire, *IRD*:

> IRD: Rationality requires that: [If you desire E, and believe M is a means to E, then you desire M].

To illustrate IRD, IRD says that you are required, if you desire beer, and believe you can get some at the pub, to desire to go to the pub. Something like IRD seems plausible: It seems perfectly rational to want things that promote your goals. Still, there are various worries for IRD. To those we now turn.

3 Problem 1: Bootstrapping

A first worry for IRD appeals to cases where the agent's pre-existing desires are themselves bad or irrational in some way. For example, imagine you desire to harm your spouse, or harm yourself. Are you then required to desire the relevant means to these ends? It might seem ludicrous to suppose you might be *irrational* for failing to desire effective means to self-harm.

However, such thoughts are no real threat to IRD, for reasons that are familiar from the existing discussion of the instrumental requirement on intention. First, one simple reply to this worry is to point out that we are concerned here only with *structural* rationality, rather than substantive rationality. Given that, it is not so obviously

[2]Somewhat relatedly, Daniel Fogal has argued for a mildly revisionary picture of structural rationality, on which it consists in competing *pro tanto* rational pressures rather than strict exceptionless requirements (Fogal 2019). Likewise, the existence of structural irrationality in desire might affect our evaluation of that revisionary project. In this paper, I set Fogal's views aside and assume the orthodox picture on which we aim to formulate strict principles of rationality rather than pro tanto pressures, and treat it as an interesting further question how Fogal's view might handle these claims about desire.

https://doi.org/10.1017/S0953820825100101 Published online by Cambridge University Press

implausible to say that you might be irrational for failing to desire effective means to self-harm: we might agree that you substantively ought not do so, and yet still think there is a failing of consistency on your part in failing to have this desire.

There is also a second reply to this objection. IRD – like parallel the requirement on intention – is best understood as stating a *wide-scope* requirement of rationality, so that it says *not* that once you have a desire for an end, that you are required [to desire the means], but instead that you are required [to desire the means to your ends] (see e.g. Broome 2004; 2013, chap. 8; for helpful discussion see Kiesewetter 2017, sec. 3.1; Worsnip 2021, sec. 6.2). One way to comply with that wide-scope requirement is to come to desire the means to your ends, but another way to comply with it is to abandon your desires for those ends. That is, assume that you do indeed desire to harm yourself, and believe that you can do so by jumping out of a window. It is true that you can become instrumentally rational by coming to desire to jump out of the window, but you can also become instrumentally rational by ceasing to desire to harm yourself. In this way, IRD does not require you to desire foolish things just because you have foolish ends; it requires only that you find *some* way to render your desires for means and ends coherent. And that is possible without desiring anything foolish.

The points here are exactly parallel to what has been said about the instrumental requirement on intention. No doubt there is more that could be said about this objection and especially about the wide-scoping reply (see e.g. Kiesewetter 2017, chaps 4, 6; Worsnip 2021, chap. 6). But since these replies are both prima facie plausible, and the residual issues perfectly general, we will focus on other issues that are unique to the case of desire.

4 Problem 2: Scepticism about instrumental desires

A second worry about IRD is that it presupposes that one can form instrumental desires. But there is a surprisingly forceful picture on which there are really no instrumental desires at all, and so no literal sense to the idea that one might be required to form them. This position may sound surprising, but is endorsed by numerous authors such as Michael Smith, Nomy Arpaly, Tim Schroeder, and Neil Sinhababu:

> "Instrumental desires are not distinct from the non-instrumental desires and means-end beliefs that explain them, but are rather just the complex state of having such non-instrumental desires and means-end beliefs standing in a suitable relation." (Smith 2004, 96)

> "instrumental . . . desires do not appear to have interesting lives of their own; they are *mere* manifestations of a person's intrinsic desires and beliefs" (Arpaly and Schroeder 2014, 9, their emphasis)

> "the combinations of intrinsic desires with means-end beliefs that are typically called 'instrumental desires' shouldn't be counted as desires" (Sinhababu 2017, 27)

Similar claims are also endorsed by David Chan (2004), and Ryan Cox (2022). On this *reductive* view, talk of instrumental desire is really just shorthand talk about the combinations of non-instrumental desires (desires!) and beliefs that people have; instrumental desires are not further attitudes in their own right. On this reductive view,

https://doi.org/10.1017/S0953820825100101 Published online by Cambridge University Press

there is never any requirement to form instrumental desires because they are not independent attitudes that one might form at all.

A small bit of terminology will be helpful: An instrumental desire has another desire and means-end belief as its *base*. The reductive view above reduces instrumental desires to their *ultimate* bases – some *ultimate desire* and a *means-end belief*.

I will highlight two problems with the reductive view above. The first appeals to thoughts about the retrieval of our instrumental desires. If the reductive view were right, when you mentally retrieve an instrumental desire, you are simply retrieving the relevant base states. But such claims are implausible. As I go to do my grocery shopping, I have in mind a set of items that I want: eggs, milk, honey, and so on – these are things that I instrumentally want. But when I bring these desires to mind, I need not bring their bases to mind – especially not their ultimate bases. I might well remember that I want milk without remembering exactly why I wanted it; I memorise the list of items precisely because that is a convenient summary of the important information for my shop, and it would be downright inconvenient if I could only recall what I wanted by also recalling exactly what I wanted those things *for*. On a more plausible picture, one's instrumental desires are further states of mind, normally causally downstream from the relevant base desires and beliefs, but nonetheless states of mind in their own right. A helpful comparison here is with belief: in principle some creature might refrain from ever believing the consequents of conditionals, instead just deploying belief in the conditional and the antecedent together in the relevant contexts. But for creatures like us – and perhaps for just about any rational creature – it makes more sense to come to believe the consequent of that conditional as a further attitude, so that you can recall it independently of its grounds. The same applies in the case of instrumental desire, where it makes perfectly good sense to store the instrumental desires you form as separate attitudes, accessible independently of their bases.

This argument is not decisive. Defenders of the reductive view may say that intrinsic desire and means-ends belief involve certain dispositions, and those dispositions might include a disposition to attend to just the relevant means in contexts where you are deciding what to do. So perhaps their view is consistent with the idea that you can retrieve the items on your shopping list, without thinking of what they are for, even though they aren't themselves objects of some new and distinct attitude of instrumental desire. So far as I can see, this is indeed a possibility. But it is somewhat ad hoc as a hypothesis. Again, think of the comparison with beliefs in conditionals: we *could* maintain that people never believe the consequents of conditionals, and what looks like retrieval of such consequents is really a dispositional effect of the underlying beliefs in the conditional and antecedent. But such a view, though possible, seems overcomplex. Why not just say that we form a separate attitude and just retrieve it directly? So the argument above, whilst not decisive, nonetheless puts some significant pressure on the reductive view.

The second problem for the reductive view aims to be more decisive, but also requires a little more discussion. To begin, note Smith's qualification that to have an instrumental desire, one must have the relevant basing desire and belief "standing in a suitable relation." This qualification serves to render the view consistent with the possibility that one might have the relevant basing desire and belief, but fail to put them together in one's mind. Perhaps, for example, you'd like to increase sales of your widgets, and learn some new information that, when added to the stock of information you already have, implies that you can increase your sales by adjusting the layout of your website. You might immediately come to desire to adjust your website, but if the relevant chain of

https://doi.org/10.1017/S0953820825100101 Published online by Cambridge University Press

inference is long or obscure enough, or you are tired, or your attention is directed elsewhere, then you might fail to form this desire. So it is surely true that one can have the basing desire and belief without having the relevant instrumental desire. Smith accepts that; hence his qualification.

But the qualification would not help if it were possible to have an instrumental desire without having the relevant basing attitudes at all. I shall argue that this is indeed a possibility (pace Arpaly and Schroeder 2014, 9; Smith 2004, 96; cf. McDaniel and Bradley 2008, 286–91). I shall argue by appeal to cases where you have an instrumental desire, then drop the basing belief, but fail to drop the dependent instrumental desire. I think other cases are possible too – where you retain an instrumental desire even after dropping the basing ultimate *desire*, or where you *form* an instrumental desire with an inadequate base – but those possibilities raise additional complications, and the case I focus on is enough by itself to undermine the reductive view.

Here is the kind of case I have in mind: Imagine that I want pleasure. I believe that *CoffeeWorld* sell pleasant coffee, and so come to desire to visit CoffeeWorld next week. But later, my friend informs me that CoffeeWorld sell only awful coffee. If I am rational, I should update my desire to visit CoffeeWorld. But I need not be rational: perhaps I take in my friend's testimony only as a general point of interest, not as something relevant to my desires for next week. If you ask me: "Do CoffeeWorld sell pleasant coffee?", I will recall my friend's testimony, and respond negatively. But if you ask me: "Do you want to visit CoffeeWorld next week?", I will nonetheless respond positively. Of course, if you ask me these questions together, I might notice the discrepancy between my attitudes and revise one. But if the issues don't come up simultaneously, my irrationality can persist. But the reductive view says otherwise: it says that as soon as I get the testimony from my friend, I necessarily and immediately cease desiring to go to CoffeeWorld. But that is not plausible; updating our desires in light of new information takes mental work.

The case seems realistic. And indeed, it seems like a common kind of case: as Smith's qualification above already implicitly concedes, changing your instrumental desires in line with changes in your other beliefs and desires is not trivial, and with that in mind it seems that one kind of predictable failing is where you fail to abandon some instrumental desire that you ought to abandon. Indeed, this is just what IRD tells us: it is a rational requirement, not an inevitability, that your instrumental desires will line up in the right way with your more fundamental desires and beliefs.

What can defenders of the reductive view say in response? So far as I can see, they have two main responses. Their first option is to say that in the CoffeeWorld case, my desire to visit CoffeeWorld has become an ultimate desire, and that is how I might continue to have this desire despite lacking any belief that the visit will bring me pleasure. In other contexts, it is familiar that instrumental desires can become ultimate desires over time: you might desire to do ballet only with the aim of impressing a love interest, but in time come to value it for its own sake. Perhaps something similar has happened in the CoffeeWorld case?

But we can see that the persisting desire to visit CoffeeWorld in such a case is plausibly not an ultimate desire by realising precisely that it will ordinarily be modified immediately if the agent *does* notice that it is unsupported by their beliefs. That is, if you make me remember that I only wanted to visit CoffeeWorld for the coffee, which I now agree is terrible, I am highly likely to abandon that desire to visit CoffeeWorld. But this is not what we would expect if that desire has somehow transformed into an ultimate desire. So it is untenable for the defender of the reductive view to insist that the CoffeeWorld case should be understood as involving a persisting ultimate desire.

https://doi.org/10.1017/S0953820825100101 Published online by Cambridge University Press

The second way in which defenders of the reductive view might reply to the CoffeeWorld case is to say that what remains, once I abandon my belief, is really only an intention, not an instrumental desire. After all, it is widely accepted that a characteristic feature of intentions is precisely that they are resistant to reconsideration, even as the agent's other relevant attitudes change (Bratman 1987). This kind of commitment is exactly what enables intentions to play useful roles in our lives, such as in coordinating behaviour with others. With this thought in mind, perhaps the CoffeeWorld case is just one more illustration of this stability of intention, showing nothing about desire at all.

For this response to be plausible, it has to be plausible that in the CoffeeWorld case I end with *only* an intention to visit CoffeeWorld, and not also a desire. But this is not plausible. For one thing, arguably intention requires desire; one cannot intend to do something if one has literally no desire to do it. If this is true, then the CoffeeWorld case simply can't be a case where an intention survives change in belief without also being a case where an instrumental desire survives change in belief. A second problem is that we could construct cases like the CoffeeWorld case where the relevant attitude is clearly not an intention, but instead a mere wish. For example, imagine that CoffeeWorld is in a distant country, one I will never visit. I might initially dream of going to CoffeeWorld, a kind of fantasy about how lovely that would be, given their extraordinary coffee. But your later testimony crushes my fantasy; it turns out that the great reputation of CoffeeWorld is a mere marketing ploy by the distant tourism board. In such a case, I might have never intended to visit CoffeeWorld given its distance. But I might instrumentally desire to go, and might continue to instrumentally desire to do so even after learning the facts. Again this is irrational, but possible. And in this version of the case, it's clear that what remains despite my change in belief is not an intention, but instead (just) a desire.

In short, in the CoffeeWorld case, it is not plausible that what persists is really just an ultimate desire, or an intention. More plausibly, it is an instrumental desire, just as I initially said. And this means that reductive theories of instrumental desire are false: we cannot understand instrumental desires as *merely* consisting in the relevant underlying desires and means-ends beliefs, not even those standing in a suitable relation. We should instead see instrumental desires as independent states of mind in their own right – as further evidenced by the facts about how we retrieve such attitudes from memory. These thoughts together serve to save IRD from the objection that such a requirement would be somehow redundant or even confused: IRD instead states a perfectly sensible requirement on our instrumental desires, which are best understood as real and independent states of mind.

5 Problem 3: Clutter

There is a third objection to IRD. Imagine someone who has rational intentions – whose intentions are all perfectly consistent, well-supported, and so on. Such a person is doing well. Is it really plausible that they might yet be irrational depending on their desires? How could that matter if their intentions were nonetheless perfectly rational (cf. Bratman 1987, 18–27; Chan 2004)? We might think that since our intentions ultimately guide our actions, our desires matter only insofar as they bear on our intentions.

As stated, this objection is implausibly strong. There are surely plenty of cases where certain combinations of desires are irrational, independently of any further facts about the agent's intentions. For example, someone who on balance desires A more than B,

https://doi.org/10.1017/S0953820825100101 Published online by Cambridge University Press

B more than C, and C more than A seems obviously irrational, and seems irrational even if this makes no difference to their intentions. For example, further imagine that they also prefer D to all of A, B, and C, so that their intransitive preferences between A-C never matter in practice. Even so, their underlying attitudes are irrational. And such irrationality in desire is not limited to only intransitive preferences – it clearly extends to various failures of instrumental rationality. For example, imagine that Sarah wants to avoid extreme unnecessary pain. We reveal that we will shortly be taking her – against her will – to the dentist for an unnecessary tooth extraction. Surely, if Sarah is rational, she will desire that we leave her alone: this is an effective means to something she wants. In such a case, the rational pressure to have such a desire is clearly independent of her intentions, since she has no choice in the matter. All the same, if she has no desire that we leave her alone – if she is indifferent to our mean behaviour – she seems irrational. Other similar cases are likewise plausible: surely rational agents want better governments, more money, better health, as well as fewer personal disasters, failures, and pains. The pressure to have such desires is generated merely by their obvious consequences for the things we care about, and goes beyond whatever pressure there is to have rational intentions about specific choices we might face.

Is there yet some grain of truth in the objection being offered? I suggest there is. Even if there are some requirements on our instrumental desires, we might think that a requirement to desire *every* means to *all* our ends seems to require too much. For example, imagine that you have some *very* mild desire to have shorter fingernails (they are marginally long; though not particularly so). You know that one way to make them shorter is to cut them with a chainsaw. Are you thereby rationally required to desire to do so? Even if we agree that there is genuinely a (surprising) thing to be said in favour of this extreme course of action, we might reasonably deny that you are rationally required to have a desire for a possibility that you will never realistically pursue. So though our initial objection was overstated, a shadow of it rings true: some absences of instrumental desire do seem perfectly rational. On one way of seeing things, the initial objection above sounded plausible only because this possibility is particularly salient in some cases where those desires make little practical difference.

We should be careful not to overstate even this more modest objection. As the fingernail case shows, we are often happy to ascribe desires even when they are extremely weak and make little difference to our conscious lives. As discussed at the outset, the term "desire" may suggest that the attitude under discussion is some uniquely intense passion, but in the relevant sense here, we have very many desires, including very many weak ones. So the objection is not merely that IRD implies that rational agents should have very many desires, including very weak desires: that should be accepted anyway. The problem is somewhat narrower: just that a requirement to desire *every* means to any of our ends seems too demanding.

In the face of this objection, we should concede that coherence in instrumental desire is sometimes unimportant. This is just as things are with the requirement to believe the apparent consequences of our beliefs: such a requirement would be too demanding if it required us to believe *every* apparent consequence of every belief we have (Harman 1986, 12). Just as the right response to that worry is to place some restriction on exactly which new beliefs we are rationally required to form (e.g. Broome 2013, 157–58), so too the solution here is to place some restriction on exactly which instrumental desires we are required to form. That is, we should endorse some version of the following view:

https://doi.org/10.1017/S0953820825100101 Published online by Cambridge University Press

> IRD-C: Rationality requires that: [If you desire E, and believe M is a means to E, then you desire M], *though only on the condition that the desire for M is not clutter*.

Assuming we can give an adequate account of which instrumental desires are "clutter," IRD-C promises to help explain why we are not required to instrumentally desire every means to every one of our ends. It thereby promises to help provide a suitably restricted version of IRD – one that rebuts the worry that an instrumental requirement on desire, if understood too broadly, would be too demanding.

What account can we give about which instrumental desires count as clutter? One option would be to restrict IRD so that it generates requirements to form instrumental desires only if those desires might somehow later bear on our intentions. This would constrain the scope of IRD so that the burdens it places on us are weaker, and help address the objection. But in fact this suggestion is over-restrictive: it seems that we can sometimes be required to form instrumental desires even in cases where those desires do not make a difference to our future intentions – consider cases like Sarah's above, where it seems that salient instrumental desires can be required even if those desires make no difference to what we might reasonably intend. More generally, it seems that we can be required to hold instrumental rational desires even about possibilities that are out of our control, or in the past, or that are otherwise insignificant with respect to our intentions. Our initial worry was that IRD was too broad, and needs constraining, but this proposal would over-constrain it, underselling the scope of the instrumental requirement.

A better option is to say that we are required to form instrumental desires only in those cases where we have attended to the relevant implication (for a similar proposal regarding closure requirements on belief, see Wedgwood 2002). We can then reformulate IRD-C as follows:

> IRD2: Rationality requires that: [If you desire E, and believe M is a means to E, then you desire M], though only on the condition that you attend to the fact that M is a means to something you desire.

IRD2 says that the instrumental requirement on desire only kicks in with respect to instrumental desires where you attend to the relevant facts, putting together your desire for the end with your belief about this means to that end. Some desires will fail this condition, as with the possibility of my desiring to cut your fingernails with a chainsaw: at least ordinarily few people will attend to the fact that they can cut their nails, as desired, by using a chainsaw, and to that extent won't be rationally required to instrumentally desire to do so. True, if they *do* attend to the way in which this would lead to something they desire, they would then be rationally required to form this desire. But that, I submit, is plausible: once you notice this possibility, there does seem to be some modest rational pressure to desire it, even if only very mildly.

Let me summarise. IRD2 promises to solve the objection above. It allows us to maintain an instrumental requirement on desire while allowing that we are not always required to desire outlandish means to minor ends of ours: it requires us only to form instrumental desires where we attend to the relevant implications for what we ultimately want. For example, IRD2 does not ordinarily require you to form desires about chopping your nails with a chainsaw, because the fact that this is a means to something you want will rarely be salient to you.

https://doi.org/10.1017/S0953820825100101 Published online by Cambridge University Press

Before we move on, I should clarify IRD2 in an important respect. We should distinguish two more careful formulations of IRD2. Remember that principles of structural rationality are often understood to be *wide-scope*, so that what is required is not some single attitude, where this requirement obtains only given some condition, but instead what is required is just to have or avoid certain combinations of attitudes: say, not to both believe P and believe ¬P (see section 3, above). As applied to IRD – the earlier version – we can helpfully restate the view as requiring you to not do all three of: desiring E, believing M is a means to E, and failing to desire M (i.e. □¬[De&Bme&¬Dm]). How about IRD2? One view might say that you are required not to do all *four* of: desiring E, believing M is a means to E, attending to the fact that M is a means to something you desire, and failing to desire M (i.e. □¬[De&Bme&Am&¬Dm]). But this is not the best version of IRD2: it implausibly says that you can satisfy the requirement by ceasing to attend to M. We should instead understand IRD2 to state a condition of the rational requirement applying at all, so that the condition that you attend to M is an external condition on the existence of the original wide scope requirement (for related discussion, see Broome 2013, sec. 8.1; and also Way 2012). That is, IRD2 effectively says that if you attend to M, then you are rationally required to obey IRD (i.e. Am→□¬[De&Bme&¬Dm]).

6 Problems 4 and 5: Immoral means, counterproductive means

Here are two further kinds of case that might cast doubt on IRD2:

First, cases where the relevant means are clearly immoral. For example, imagine that I desire a larger house, and attend to the fact that killing my father for his money would help me afford the house (perhaps my immoral wife asserts this, pointedly). Might IRD imply that I am rationally required to desire my father's death? Note that the problem here is independent of the bootstrapping problem canvassed in section 3. There the problem was generated by ends that were themselves inappropriate. But the problem here is generated even in cases where your ends are perfectly appropriate: there is nothing wrong with wanting a larger house. But does this legitimate end rationally require desires for just *any* way of achieving it?[3]

Second, cases where the relevant means is clearly counterproductive to your ultimate desires. For example, imagine that I ultimately desire happiness, and come to instrumentally desire money as a means to that end. I might pursue *that* means by pursuing a lucrative but miserable career. Am I therefore rationally required to desire this miserable career, as a means to satisfying my instrumental desire for money? Surely not if it worsens, rather than improves, my chances of satisfying my original ultimate desire for happiness. Does IRD2 secure that result?

My reply to these worries stems from a broader strategy for defending IRD2. That strategy is to note parallels between IRD and questions about the instrumental transmission of normative reasons (for some issues and references, see Kolodny 2018; Kiesewetter and Gertken 2021). We can respond to the above objections to IRD2 by borrowing from existing views about the instrumental transmission of reasons. One neat feature of this strategy is that even if the specific proposals below later turn out to be problematic, the basic strategy should survive: *whatever* the correct view about the

[3]Needless to say, the move to IRD2 does not help with this problem, since you might attend to this very implication as in the case above. In what follows I largely set aside the extra condition imposed with IRD2, since it fails to address cases of this kind.

https://doi.org/10.1017/S0953820825100101 Published online by Cambridge University Press

instrumental transmission of reasons, one might reasonably think that we could borrow from it to construct the right materials to rescue IRD2 from these objections.

With respect to the transmission of instrumental reasons to even immoral means, one relatively standard view on that topic is to insist that such reasons really do exist; that reasons really are transmitted from legitimate ends to immoral means to those ends – perhaps you really do have some reason to murder your father. Such claims sound surprising, but their counterintuitiveness has a natural explanation, which is that since the relevant reasons are so massively outweighed by countervailing reasons (e.g. the exceptionally strong reasons *not* to murder your own father), it is often inappropriate to think about such reasons, and inappropriate even to assert their existence, which might conversationally imply that they have some practical relevance (for critical discussion and further references, see Kiesewetter and Gertken (2021); see also Schroeder (2007, chap. 5)).

The parallel claim in defence of IRD2 is to insist that we are rationally required to desire the relevant means even if those means are immoral. But we can then add that the desires for those immoral means should be incredibly weak relative to other desires *not* to take such immoral courses of action, and that it is this relative weakness that makes it potentially misleading to talk about requirements to form such desires (cf. Gregory 2021, sec. 2.4.1, 5.4). That is, perhaps you are rationally required to desire your father's death, for the sake of the inheritance. But asserting this may well be inappropriate since the desire in question ought to be extraordinarily weak and irrelevant for all practical purposes; in many contexts mentioning this requirement may suggest that the desire matters in some way. Since it does not, it is in many contexts inappropriate to mention this requirement.

The other objection above was about counterproductive means. Again, existing discussion about the instrumental transmission of reasons provides the right materials for responding to the objection. There, the best view says that only *non-instrumental* reasons can generate new instrumental reasons; one cannot iterate instrumental transmission and generate reasons to pursue means *to the means* to compliance with ultimate reasons (e.g. Bedke 2009, 679n12). So with respect to reasons to pursue miserable careers, you might have a reason to pursue happiness, and this might give you a reason to pursue money. But the latter reason, being instrumental, cannot itself generate further instrumental reasons. So you may have no reason to pursue lucrative but miserable jobs, since such reasons could be generated only by the ultimate reason to pursue happiness, not the instrumental reason to pursue money. Likewise, with respect to IRD2, the view should be clarified as follows:

> IRD3: Rationality requires that: [If you *ultimately* desire E, and believe M is a means to E, then you desire M], though only on the condition that you attend to fact that M is a means to something you desire.

One intuitive way to put this is to say that instrumental desires are not appropriate inputs to the instrumental requirement. As a result, a desire for happiness might rationally require a desire for money, but there is no rational requirement to pursue miserable but lucrative careers: this is not a means to the ultimate desire for happiness, and there is no requirement to form desires that are only means to one's merely instrumental desires.

This move does not merely serve to save IRD from an objection; the move to IRD3 has widespread significance with respect to our capacity to be rational in our desires. As

https://doi.org/10.1017/S0953820825100101 Published online by Cambridge University Press

I said in section 4, we can retrieve our instrumental desires independently of retrieving their bases, and this is often very helpful. It is often helpful because the chain of reasoning from one's ultimate desires to some new instrumental desire might be quite long, and it might often be easier to shorten it, by forming new instrumental desires just by appeal to one's old instrumental desires. For example, you might form desires about what you need at the shop just starting with your desire – surely instrumental – to make a lasagne tonight. IRD3 doesn't exactly militate against doing this, but it does highlight the risk, which is precisely that you might sometimes form new instrumental desires that serve *only* other instrumental desires of yours, not your ultimate desires, and in those cases your new instrumental desires are not rationally required and might even be rationally impermissible, depending on your other desires. This is just the lesson from the case – all too realistic – where you end up desiring to pursue some lucrative but miserable career as a means to wealth, even though you desire wealth only as a means to happiness, which the job will frustrate. That is, IRD3 rightly tells us that instrumental desires had better serve your ultimate desires, not just other instrumental desires of yours. But pragmatic considerations often favour forming instrumental desires by appeal only to other instrumental desires of ours, and we thereby run the risk of forming counterproductive desires. This kind of irrationality is common, significant, and IRD3 is right to highlight it.

*****

The various problems and solutions above can all be illustrated nicely by an example from John Broome (2005), later endorsed by Ralph Wedgwood (2011). They present the following objection to an instrumental requirement on desire: You are hungry, and want to be rid of your hunger. You believe that killing yourself would be a way of getting rid of your hunger. They say it is implausible that you are rationally required to desire to kill yourself. But in light of the claims above, we can see that this case is unconvincing. I shall rehearse four of the points above that bear on this case.

First, given the anti-clutter qualification above, you would only be rationally required to form this desire if you happen to attend to the way in which killing yourself would help alleviate your hunger. If the objection is simply that a requirement in such a case would be too demanding given that this means to this end is not at all salient, the anti-clutter condition introduced in Section 5 accommodates it perfectly.

Second, even if you do consider this means to alleviating your hunger, the desire you are required to form would be very weak – easily outweighed by various countervailing desires – and to that extent it is not so obviously counterintuitive to suppose that you might, on considering the matter, find there is some rational pressure to form this weak desire (again, cf. Gregory 2021, sec. 2.4.1, 5.4).

Third, as argued in Section 3, structural norms of rationality are wide-scope norms. So a further possibility in this case is that one should comply with IRD3 by abandoning the desire to be rid of your hunger. Perhaps this desire sounds reasonable only because it is easily confused with nearby and more appropriate desires, such as the desire to be full. That desire would not be served by suicide, and so perhaps Broome and Wedgwood's example gets its force in part because it appeals to a desire that rationality forbids and so where instrumental rationality should consist in abandoning that end, not desiring the relevant means.

Fourth, as argued in this section, only ultimate desires can rationally require further instrumental desires. This provides another reason why there is no requirement to desire to kill oneself in the case in question. After all, the desire to be rid of your hunger is

https://doi.org/10.1017/S0953820825100101 Published online by Cambridge University Press

plausibly an instrumental desire: perhaps hunger involves a positive ultimate desire [for food], but the more complex desire [to be rid of hunger] is clearly derivative of that or some other desire you have. But given IRD3, a merely instrumental desire to be rid of hunger generates no rational pressure to form further desires. An ultimate desire for food could create rational pressure to form certain instrumental desires, of course, but killing yourself is not a means to *that* end.

In short, Broome and Wedgwood's example can be handled in numerous ways and on reflection provides no evidence at all against the existence of an instrumental requirement on desire. IRD3 remains a plausible requirement on desire.

7 Two further points of detail

Before we move on from the instrumental requirement, I should briefly note two remaining questions about it.

First, we might enhance IRD3 with claims about the required *strength* of the relevant instrumental desires (cf. Fogal 2019). For example, we might think that we are rationally required to have instrumental desires whose strengths are proportionate, in the obvious way, to the strengths of our non-instrumental desires weighted by our credences in the relevant means-ends beliefs. Such a view is highly natural but there are respectable views on which other attitudes to risk are rationally permissible (see especially Buchak 2013). I leave open exactly how IRD should be formulated in light of these issues.

Second, we might want to supplement IRD with claims about how desires for means aggregate together in strength (or don't). For example, imagine that I ultimately desire both E and F, and M is a means to both. Should I desire M more than if it were just a means to E? It seems in some cases that the answer is positive – a food that delivers both taste and health might be more desirable than one that delivers just one of those things. But in other cases, the answer might be negative – a food might deliver two tastes which are individually desirable but not in combination. Again, I won't try to resolve this issue here, though I again note that similar issues arise about how reasons aggregate (see e.g. Maguire and Snedegar 2020), and we might again borrow answers from that existing discussion.

In short, these issues are interesting, but I won't try to resolve them here.

8 Other rational requirements on desire

I conclude that IRD3 is a genuine rational requirement on desire. Is it the only rational requirement on desire? No. There are other rational requirements on desire that connect them to our normative beliefs and our intentions. In a fully coherent agent – a fully structurally rational one – their normative beliefs, desires, and intentions, will all align neatly.

That is, we should endorse the following requirements regarding how our desires align with our normative beliefs:

a) Rationality requires that: [If you believe you have reason to E, then you desire to E].
b) Rationality requires that: [If you believe you ought to E, then you desire to E more strongly than any available alternative to E].

Requirements like these say that our desires ought to reflect our wider views about what is worth pursuing. Various qualifications of these requirements are possible – for

https://doi.org/10.1017/S0953820825100101 Published online by Cambridge University Press

example, they could be strengthened to biconditionals, so that the relevant desires are *only* rationally permissible if you have the requisite beliefs. Still, although there are interesting points of detail to discuss, it seems highly plausible that there are rational requirements along these lines.

Likewise, we should endorse the following requirements regarding how our desires align with our intentions (see also Gregory 2021, chap. 4 and especially p78):

c) Rationality requires that: [If you most strongly desire to E, and believe you can E, then you intend to E].
d) Rationality requires that: [You must: If you most strongly desire to E, believe you can E, and believe E→¬F, then you don't intend F].

Again, various qualifications of these requirements are possible: we might strengthen them to biconditionals saying also that the relevant intentions are rationally permissible *only* if you have the requisite desires. Other modifications are also possible – most importantly, (c) needs amending to permit that you might rationally lack the intention in cases where it would be redundant, or where you might reasonably instead adopt it later in time (for discussion see e.g. Broome 2013, 170–72; Kiesewetter 2017, 191–92). Again, although there are interesting points of detail to discuss, it is highly plausible that there are rational requirements along these lines.

As stated at the outset, the existence of these requirements – and within limits, their exact formulation – may have widespread significance. One possible application I mentioned was the viability of a middle way between reasons internalism and reasons externalism. A requirement such as (c) above would help do that, by capturing some of the appeal of reasons internalism, treating the crucial part of the view as the idea that rational intention hinges on our desires. The requirements connecting desire to intention also have additional interest. When discussing rationality in intention, many authors standardly discuss the *Enkratic requirement* to intend to do the things you believe you ought to do, as well as the requirement to have consistent intentions. Such requirements might derive from upstream constraints on desire, as given in (b)-(d) above. For example, perhaps the Enkratic requirement obtains just because you are rationally required to most strongly desire to do whatever you believe you ought to do, and also rationally required to intend to do whatever you most strongly desire to do. This need not undercut the Enkratic requirement on intention, nor render it redundant, but it might nonetheless provide a deeper vantage point from which to understand the nature and grounds of that requirement.

9 Conclusion

I have argued for structural requirements of rationality on desire. In most depth I defended the instrumental requirement on desire, showing how that constraint can escape various objections – from cases where one's ultimate desires are bad, from worries about the very existence of instrumental desires, from worries about desires with no practical relevance, from worries about desires for immoral means to one's ends, and from worries about desires for means that are ultimately counterproductive. I also more briefly endorsed some further structural requirements on desire. All these requirements have intrinsic interest, but also have wider significance in metaethics.

https://doi.org/10.1017/S0953820825100101 Published online by Cambridge University Press

Acknowledgements. For helpful comments on this paper, I thank audiences at LSE, Tartu, and Southampton, and especially, Jonathan Way, and two referees for this journal.

References

Arpaly, Nomy and Timothy Schroeder, *In Praise of Desire* (New York: Oxford University Press, 2014).
Bedke, Matthew S, 'The Iffiest Oughts: A Guise of Reasons Account of End-Given Conditionals', *Ethics* **119** (2009), 672–98. https://doi.org/10.1086/600130.
Blackburn, Simon, *Ruling Passions: A Theory of Practical Reasoning* (New York: Oxford University Press, 1998).
Bratman, Michael, *Intention, Plans, and Practical Reason* (Cambridge, Mass: CSLI Publications, 1987).
Broome, John, 'Reasons', in *Reason and Value: Themes From the Moral Philosophy of Joseph Raz*, ed. by R. Jay Wallace (Oxford: Oxford University Press, 2004), pp. 2004–28.
Broome, John, 'Have We Reason to Do as Rationality Requires? - A Comment on Raz', *Journal of Ethics and Social Philosophy* **1** (2005), 1–10. https://doi.org/10.26556/jesp.v1i1.13.
Broome, John, *Rationality through Reasoning* (Oxford: Wiley Blackwell, 2013).
Brunero, John, *Instrumental Rationality: The Normativity of Means-Ends Coherence* (New York: OUP Oxford, 2020).
Buchak, Lara, *Risk and Rationality* (New York: Oxford University Press, 2013).
Chan, David K, 'Are There Extrinsic Desires?', *Noûs* **38** (2004), 326–50.
Cox, Ryan, 'Constitutivism About Instrumental Desires and Introspective Belief', *Dialectica* **74**(4) (2020), 635–64. https://doi.org/10.48106/dial.v74.i4.02.
Fogal, Daniel, 'Rational Requirements and the Primacy of Pressure', *Mind* **129** (2019), 1033–70.
Fogal, Daniel, and Alex Worsnip. 2021. '*Which Reasons? Which Rationality?*' *Ergo* **8** (11). https://doi.org/10.3998/ergo.1148.
Fullhart, Samuel, and Camilo Martinez, 'Coherence as Joint Satisfiability', *Australasian Journal of Philosophy* **102** (2024), 312–32. https://doi.org/10.1080/00048402.2023.2269963.
Gregory, Alex, *Desire as Belief: A Study of Desire, Motivation, and Rationality* (Oxford: Oxford University Press, 2021).
Harman, Gilbert, *Change in View: Principles of Reasoning* (Cambridge: The MIT Press, 1986).
Heathwood, Chris, 'Which Desires Are Relevant to Well-Being?', *Noûs* **53** (2019), 664–88. https://doi.org/10.1111/nous.12232.
Hume, David, 1738. *A Treatise of Human Nature*.
Kiesewetter, Benjamin, *The Normativity of Rationality* (Oxford: Oxford University Press, 2017).
Kiesewetter, Benjamin, and Jan Gertken, 'How Do Reasons Transmit to Non-Necessary Means?', *Australasian Journal of Philosophy* **99** (2021), 271–85. https://doi.org/10.1080/00048402.2020.1745252.
Kolodny, Niko, 'Why Be Rational?', *Mind* **114** (2005), 509–63.
Kolodny, Niko, 'Instrumental Reasons', in *The Oxford Handbook of Reasons and Normativity*, ed. by Daniel Star (Oxford: Oxford University Press, 2018).
Lauer, Sven, and Cleo Condoravdi, 'Preference-Conditioned Necessities: Detachment and Practical Reasoning', *Pacific Philosophical Quarterly* **95** (2014), 584–621. https://doi.org/10.1111/papq.12049.
Lewis, David, 'Desire as Belief', *Mind* **97** (1988), 323–32.
Lord, Errol, *The Importance of Being Rational* (Oxford: Oxford University Press, 2018).
Maguire, Barry, and Justin Snedegar, 'Normative Metaphysics for Accountants', *Philosophical Studies* **178** (2020), 363–84. https://doi.org/10.1007/s11098-020-01435-w.
Markovits, Julia, *Moral Reason* (Oxford: Oxford University Press, 2014).
McDaniel, Kris, and Ben Bradley. 'Desires', *Mind* **117** (2008), 267–302. https://doi.org/10.1093/mind/fzn044.
Nagel, Thomas, *The Possibility of Altruism* (Princeton, N J: Princeton University Press, 1970).
Parfit, Derek, *On What Matters: VoluOn What Matters: Volume I* (New York: Oxford University Press, 2011).
Scanlon, T. M, *What We Owe to Each Other* (Cambridge, Mass.; London: Belknap, 1998).

Scanlon, T. M, 'Structural Irrationality', in *Common Minds: Themes from the Philosophy of Philip Pettit*, ed. by Geoffrey Brennan, Michael Smith, Robert Goodin, and Frank Jackson (Oxford University Press, 2007), pp. 84–103.

Schroeder, Mark, *Slaves of the Passions* (Oxford; New York: Oxford University Press, 2007).

Schueler, George Frederick, *Desire: Its Role in Practical Reason and the Explanation of Action* (Cambridge, Mass.; London: MIT Press, 1995).

Sinhababu, Neil, *Humean Nature: How Desire Explains Action, Thought, and Feeling* (Oxford: Oxford University Press, 2017).

Smith, Michael, *The Moral Problem* (Oxford: Oxford University Press, 1994).

Smith, Michael, 'Instrumental Desires, Instrumental Rationality', *Proceedings of the Aristotelian Society, Supplementary Volumes* **78** (2004),93–109.

De Sousa, Ronald B, '*The Good and the True*', *Mind* **83** (1974), 534–51.

Steele, Katie, and H. Orri Stefánsson, 'Decision Theory', in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta, Winter (Metaphysics Research Lab, Stanford University, 2020). https://plato.stanford.edu/archives/win2020/entries/decision-theory/.

Way, Jonathan, 'Explaining the Instrumental Principle', *Australasian Journal of Philosophy* **90** (2012), 487–506. https://doi.org/10.1080/00048402.2011.598174.

Wedgwood, Ralph, 'The Aim of Belief', *Philosophical Perspectives* **16** (2002),267–97.

Wedgwood, Ralph, 'Instrumental Rationality', *Oxford Studies in Metaethics* **6** (2011),280–309.

Williams, Bernard, 'Ethical Consistency', in *Problems of the Self: Philosophical Papers 1956-1972* (Cambridge: Cambridge University Press, 1976), pp. 166–86.

Williams, Bernard, 'Internal and External Reasons', in *Moral Luck: Philosophical Papers 1973-1980* (Cambridge: Cambridge University Press, 1981), pp. 101–13.

Williams, Bernard, 'Internal Reasons and the Obscurity of Blame', in *Making Sense of Humanity: And Other Philosophical Papers 1982-1993* (Cambridge: Cambridge University Press, 1995), pp. 35–45.

Worsnip, Alex, *Fitting Things Together: Coherence and the Demands of Structural Rationality* (Oxford, New York: Oxford University Press, 2021).

Cite this article: Gregory, Alex. 2025. Structural Rationality in Desire. *Utilitas* 1–17. https://doi.org/10.1017/S0953820825100101