

Immoral realism

Max Khan Hayward¹ 

© Springer Nature B.V. 2019

Abstract Non-naturalist realists are committed to the belief, famously voiced by Parfit, that if there are no non-natural facts then nothing matters. But it is morally objectionable to conditionalise all our moral commitments on the question of whether there are non-natural facts. Non-natural facts are causally inefficacious, and so make no difference to the world of our experience. And to be a realist about such facts is to hold that they are mind-independent. It is compatible with our experiences that there are no non-natural facts, or that they are very different from what we think. As Nagel says, realism makes scepticism intelligible. So the non-naturalist must hold that you *might* be wrong that your partner (for example) matters, even if you are correct about every natural, causal fact about your history and relationship. But to hold that conditional attitude to your partner would be a moral betrayal. So believing non-naturalist realism involves doing something immoral.

Keywords Ethics · Metaethics · Moral realism · Non-naturalism · Non-naturalist realism · Moral nihilism

1 Introduction

Moral Realists think that some moral claims are true, and that's not a matter of what anyone happens to think. It's a "judgment-independent" fact. According to a prominent version of Realism, supported by Parfit, Scanlon, Nagel, Enoch and many others, it's not a *natural* fact either. True moral claims state beliefs that correspond with irreducible, purely normative facts.

✉ Max Khan Hayward
mh3173@columbia.edu

¹ Bowling Green State University / University of Sheffield, 324 Shatzel Hall, Bowling Green, OH 43403, USA

Parfit pointed out a dire implication of this view. Non-naturalists don't just think that moral claims *happen* to be made true by corresponding to non-natural facts. That's the only way that moral claims *could* be true. Other theories of moral truth *miss* something that matters. So Parfit claimed that if naturalism is true then nothing matters. There is nothing you ought to do (Parfit 2011). And this isn't just about Parfit: as I argue, it is simply implicit in the view. For the non-naturalist, naturalism entails nihilism.

For this reason, Parfit accused his philosophical opponents, such as Bernard Williams, of accepting an objectionable moral nihilism. In similar spirit, Dworkin (1996) once argued that anti-realists like Blackburn held something *immoral* when they denied objective truth to ethical claims.

Williams (2006) once said that those who accused him of "scientism" for denying ethics and aesthetics a place alongside physics in the "absolute conception" of the world were actually themselves *counterfactually scientific*—why assume ethics and aesthetics need to belong to any absolute conception in order to matter?

We can respond similarly here. The *real* problem is the claim that, if naturalism is true, then nothing matters. Why think that ethics needs non-natural, mind-independent, objective truthmakers to be authoritative? It is morally wrong to accept this conditional. It is *counterfactual* nihilism. Even if there were no non-natural truths, every other fact about, say, your partner and your relationship would be the same. It would be a betrayal to them to abandon your moral commitments just because you thought naturalism was true.

The plan is as follows. First, I outline my argument against non-natural moral realism. Realists of this stripe make us *conditionalise* our world-directed moral commitments on the wrong things. Questions about how to conditionalise our moral commitments are normative questions, and hence my objection is a normative objection. My target is primarily non-naturalism, but I'll briefly suggest that my arguments may extend to implicate other views. And then I'll work through a set of objections to the form of argument that I offer.

2 Loomings

As Dworkin's (1996) piece exemplifies, the notion that there is something morally offensive in rejecting realism is one of long standing. Indeed, it's not uncommon to hear philosophers make jokes to this effect in casual conversation—the sort of joke that belies a widespread assumption. Nevertheless, there are forebears to my inversion of the realists' accusation. Blackburn, for example, in his (1996) response to Dworkin's attack, points to the unsavory note of stridency in the exhortation that one had "better believe" that ethics admits of Objectivity and Truth. In his (1993) Blackburn argued that the non-naturalist showed a corrupt sensibility, and Bedke (forthcoming) develops the claim that there is something objectionable in the psychology of the non-naturalist. Nowell-Smith (1954) asked why we ought to care about *sui generis* properties of rightness and wrongness, and more recently, Dasgupta (2017) has asked what *authority* non-natural facts have over us. All these arguments share a common thread—they challenge non-naturalism on normative,

and in many cases, normative ethical, grounds. My argument is not identical to any of these, but it belongs in this tradition. The problem with non-naturalism is that, to the extent that one believes non-naturalism, one believes something immoral. This is because non-naturalism implies that we must conditionalise all our world-directed ethical beliefs on the existence of non-natural facts.

3 Conditionalisation

We have all sorts of moral beliefs whose objects are located in the world around us. I ought to comfort my partner who is in pain, or to keep the promise I made to my friend that I wouldn't speak too fast at my APA presentation.

To understand the totality of a person's moral outlook, it's not enough just to know their directly world-directed moral attitudes and beliefs. Our moral outlooks are revisable. There are circumstances where we think we ought to change our views. So we also have higher-order moral judgments about the conditions in which we ought to change our moral judgments.

These higher-order judgments form a major part of everyday morality. I have a special duty to my partner, for example, to comfort her in distress. But it's a familiar view that I ought to change my sense of what I owe to my partner if I discover that she has been, say, cheating on me for the past 5 years.

We frequently evaluate the correctness of one another's moral views on the basis of whether they accept the right "change conditions." You and I may have just the same commitments to our partners. But if you conditionalise these, not just on fidelity, but also on your partner maintaining a certain body mass index (BMI), then our moral viewpoints are different. Even if I also think that my partner *should* maintain a given BMI for health reasons, the fact that I don't conditionalise my other moral beliefs on this fact, and you do, is part of the reason why I am a good partner and you are a nasty brute.

Likewise, a certain kind of rule-utilitarian and Kantian may agree about what immediate, world-directed moral beliefs and attitudes are appropriate, given how the world actually is. But they will take entirely different considerations to count in favour of changing their views. The utilitarian will drop her opposition to lying if she discovers that holding a more flexible principle will deliver more utility. This is why, despite superficial agreement, the Kantian views her utilitarian colleague as believing something morally objectionable.

There are some plausible norms for widespread higher-order conditionalisation, which legislate situations where we should radically revise our moral views, or stop having any world-directed moral attitudes at all (although we would retain higher-order views). If there is no external world, and all the other people I think I interact with are mere figments of my imagination, then it seems like I should have no world-directed moral beliefs at all—at least, no other-regarding moral beliefs—and concern myself only with the requirements of self-interest or prudence. Moral nihilism—in the sense of abandoning all world-directed moral commitments—

might seem appropriate in this situation (if we accept that morality is a distinct form of normativity from prudence). But in other cases we *must not* become nihilists. Some conditionalisations are wholly reprehensible.

4 What is non-naturalist realism?

Why do I think that Non-naturalist realism forces us to conditionalise moral commitment on something objectionable? I think we can identify three general claims, in virtue of which such theories are realist, non-naturalist, and involve claims about normativity, respectively. These minimal claims will be all I rely upon in my argument.

4.1 Realism

Philosophers sometimes speak as though there were no more to the issue of Realism than the question of whether some ethical claims are true. This is unhelpful. Almost all metaethical theories allow us to call certain ethical claims “true” in some sense—constructivists, quasi-realists, even fictionalists and subjectivists all use the language of truth at various points. It would both render Realism largely useless as a term of art, and poorly follow accepted usage, if we counted constructivism, quasi-realism, fictionalism and subjectivism as forms of realism.

Realists don’t just think that some people get things right. They also say that some people—even entire communities and cultures—have got things *wrong*. Perhaps, once upon a time, the entire human race was morally mistaken. Realists have a particular way of explaining how this is possible. When people get things wrong, they’re not just wrong about what they want, or about what they *would* want if they were more sympathetic, or about what would be a good solution for their shared problems, or about what moral ideas it would be useful to invent. When we call them wrong, we’re not “merely” expressing our own critical attitudes. These people are *wrong* about the moral facts—and that’s not a matter of what anyone happens to think. These facts are “judgment-independent”.¹ Of course, by symmetry, this means that we, too, could in principle be entirely wrong about the purportedly judgment-independent moral facts. As Nagel once said “Realism makes skepticism intelligible” (Nagel 1986, p. 90).

¹ Some philosophers use “mind-independence” to characterise realism. But this gives the unfortunate suggestion that the realist cannot identify constituents of minds—such as pleasures and pains—of being the fundamental objects of normative significance. Better terms include “judgment-independence”, “stance-independence” or “response-independence,” although selecting between these terms would take further argument. Certainly, realists will want to say that at least some moral truths are independent of *all* of these things—the judgments, stances and responses of humans. For convenience, I will talk mainly of “judgment-independence.”

4.2 Non-naturalism

What makes judgment-independent facts *non-natural*? Again, the literature here is complex. Some “naturalist” realists think that there are judgment-independent moral facts, because moral concepts refer to natural facts, which are judgment-independent. These are discovered through empirical investigations.

But, non-naturalists object: couldn't some people have the *wrong* moral concepts? There has to be some deeper fact about what natural facts matter morally. And that *couldn't* be discovered through empirical investigations they say. These deeper moral facts are supposed to be *non-natural* facts. Parfit writes:

The most fundamental normative facts are not...empirically discoverable facts (Parfit 2011, Vol. II, p. 307).

Similarly, Scanlon frequently contrasts “empirical” discoveries and judgments with normative ones.

What distinguishes these two kinds of facts? Causal facts *are* empirically tractable—they are open to perceptual and experimental detection, and can enter into natural-scientific explanations. Hence almost all non-naturalists agree that irreducible normative facts must be *non-causal*. Parfit writes:

When some fact has the property of *being* or *giving us* a reason, we cannot be causally affected by this normative property (Parfit 2011, Vol. II, p. 493).

Another non-naturalist, Christian Coons, writes:

It seems like a category error to claim that facts about what we ought to do somehow explain or cause particular events that happen in the world (Coons 2011, p. 85).

This is not to say that causality is the defining mark of the natural. If there were a God, he or she would presumably have causal effects on the world (notwithstanding the old puzzle of how a perfect, unchanging, eternal being *could* cause changes in the movement of objects through space over time). But divinities are supernatural, not natural. And this portrayal can be agnostic as to whether *all* natural facts are causal or empirically discoverable. Rather, the claim is more minimally, that, so long as normative facts are understood as non-natural, they are understood as non-causal.

This may seem strange. But non-naturalists still think that non-natural, moral facts still reliably *supervene* on certain natural facts. So by tracking the relevant natural facts, we can still be responding to the non-natural, moral facts. This might sound epistemologically obscure, but realists in mathematics and other domains say similar things; so non-naturalists can, and do, appeal to a “companions in guilt” argument.²

So, when non-naturalists say that people could be entirely wrong about morality, they don't just mean that they could be wrong about the natural facts. For them, not

² The problem of giving an epistemological story for knowledge of non-causal facts is a major theme in much of the “debunking” literature, such as Benacerraf (1973), Field (1989), Street (2006) and Clarke-Doane (2012). Notable responses are given in Parfit (2011, Vol. II Ch. 32), Scanlon (2014, Ch. 4), Enoch (2011) and Clarke-Doane (2016).

only does scepticism make sense in principle, but this kind of scepticism is *independent* from any scepticism about the natural world.

4.3 Normativity

What is at stake in making these claims? Why does it *matter*? Non-naturalists don't typically think that the above claims simply describe how we *happen* to think about morality. They don't think that moral claims could, just as well, be made true just by natural facts or by our judgments and attitudes. After all, they want to claim that people can be wrong about ethics even if they're right about what they want, about all the natural facts, and so on.

So they normally suppose that *only* non-naturalism captures what is necessary for something to count as a genuinely moral truth. Scanlon writes:

To identify a reason with a naturalistic property seems immediately to destroy its normativity (Scanlon 2014, p. 46).

And Parfit makes such claims on many occasions, for example:

Normativity is either an illusion, or involves irreducibly normative facts (Parfit 2011, Vol. II, p. 267).

similarly:

If there were no such [non-natural] truths, there would be no point trying to make good decisions. Nothing would matter, and there would not be better or worse ways to live. (Parfit 2011, Vol. II, p. 425).

According to these philosophers, a discourse that simply attempted to express widely held sentiments or to identify solutions to shared practical problems would lack a special kind of *normative authority*. It is only because moral language and practice attempts to track non-natural, judgment-independent truths that it counts as genuinely normative. If theorists think that, then they should agree with Parfit's claim that if there were no truths of the relevant sort, moral talk and practice would lack genuine normative force. Nothing would matter.

5 What is non-naturalist realism: summary

We can summarise the claims of the preceding sections as follows. According to Non-naturalist Realism:

- (A) Some moral facts are independent of what anyone happens to think.
 - > Moral scepticism at least makes sense.
- (B) Moral facts are causally inert.
 - > Moral facts make no difference to the experienced world.

- (C) Only non-naturalist realism explains how things matter in the normative sense.
- > If there are no facts of the sort defined in A&B, nothing matters in the normative sense.

6 Three objectionable conditionals

What is wrong with this? As I mentioned before, we frequently criticise one another for the conditionalisations we make. It's wrong to conditionalise your commitments to your partner on him or her maintaining a certain BMI.

Furthermore, there are some conditions in which nihilism, in the sense of abandoning any world-directed moral attitudes or beliefs, makes sense. If there were no external world, or it had none of the relevant objects of moral concern, then in a moral (as opposed to prudential) sense, I should conclude that nothing would matter. But there are some cases where it seems entirely reprehensible to accept a nihilistic conclusion.

1. The Mets Fan's conditional

Some people really love the Mets. A Mets fan might exclaim:

If the Mets lose tonight, nothing matters.

Of course, no one really means this. It's just said for dramatic effect. But if anyone did, it would be morally reprehensible. The Mets losing just isn't a good reason to abandon all one's moral beliefs and attitudes.

2. Ivan's conditional

Some people think that morality needs a God to give it a special kind of authority. Ivan, in the Brothers Karamazov, voices such a sentiment:

If God is dead, then everything is permitted.

There are those who find this sentiment entirely acceptable. I do not. Even if I thought a God would matter enormously for morality, were there such a being, I should still realise that, God or no, there are real people with real needs and real pains down below.³ It scorns them if I am prepared to abandon my sense that there

³ James appears to have had such a view:

Whether a God exist, or whether no God exist, in yon blue heaven above us bent, we form at any rate an ethical republic here below. And the first reflection which this leads to is that ethics have as genuine and real a foothold in a universe where the highest consciousness is human, as in a universe where there is a God as well. "The religion of humanity" affords a basis for ethics as well as theism does. (James 1956, p. 198)

For James, God, if he exists, has a moral standing, but it is not different *in kind* from that of any human person - the response to God is simply that of "Life answering to life." James' ethics does not conditionalise moral commitment on anything that I object to.

are some things I may, and some things I must not, do to these people, just because I find out there is no God.

3. Parfit's conditional

Derek Parfit said that, if Naturalism is true, then nothing matters. I think this claim is objectionable for just the same reasons that Ivan's conditional and the Mets Fan's conditional are objectionable. And I don't think this was an eccentricity of Parfit's. Rather, I think the objectionable conditional simply follows from non-naturalist realism.

As stated in C, non-naturalists are committed to the view that, if there are no non-natural facts, then nothing matters. If naturalism is true, then nothing matters. And, as stated in A, moral realists think that moral scepticism makes sense – since moral facts are supposed to be entirely independent of what we think about morality, we could be wrong that there are any such facts. And, according to B, we could be wrong about the existence of moral facts without being wrong about the causal world that we experience. Everything could be just the same in the perceptible aspect of my relationship with my partner, our attitudes and emotions could be just the same, and yet it be false that I ought to comfort her in pain. She might not matter.

This might sound strange. The denial that anything matters is, after all, the purview of the Error Theorist, not the Realist. Realists *do* think that things matter, and that I ought to comfort my partner if she is in pain. But my point is that there is a basic structural similarity between these two positions:

CONDITIONAL PREMISE: Only if there are non-natural moral truths do some things matter.

EXISTENTIAL PREMISE: There are non-natural moral truths.

REALIST CONCLUSION: Some things matter.

Error theorists and realists disagree about the second premise and hence the conclusion. But they agree about the first. And it is this *conditional* to which I object.

Of course, most theories allow for the possibility of moral error, and will therefore conditionalise obligations to (say) my partner on *something*. But my objection is not to the very fact of conditionalisation, but to its content. As I've argued, questions about conditionalisation are naturally understood as normative ethical questions. So we can ask whether the things upon which any given theory conditionalises morality are *really* the sorts of things that need to be the case for it to be right for us to have any world-directed moral attitudes. Perhaps there are some situations where it really does make sense not to have world-directed moral attitudes, but to identify them as such is to make a moral claim.

It seems clear to me that the conditionalisation implicit in non-naturalist realism is morally unacceptable. Whether or not naturalism is true would make no causal difference to the world that we—and those we have moral relations with—live in.

There might be no non-natural moral truths, and yet the rest of our beliefs about the natural world, and about the desires and attitudes of my partner and the commitments I have made to her, could all be correct. If I were to drop my commitment to my partner, but every aspect of our history, every fact about her feelings and concerns stayed the same, that, I think, would constitute as deep a betrayal to her as abandoning my commitment because the Mets lost or because God is dead.

The case of non-naturalism is particularly vivid, because the existence (or not) of non-causal *sui generis* purely normative facts is *clearly* isolated from the things we *should* conditionalise world-directed moral commitments upon. But it seems plausible to me that similar objections might be run against other kinds of realist views, especially those that closely resemble non-naturalism. The more stringent the demands the metaethicist places upon moral truth or correctness—that it must exhibit mind-independence, objectivity, universality, it must be compelling to rationality and so on—the easier it is to object: would a world in which nothing matched this description *really* be a world in which nothing matters morally?

This concludes my direct argument. The rest of this paper returns to the case of non-naturalist realism, and focuses on objections to the argument I have just given.

7 One and the same?

My argument is that the non-naturalist must, objectionably, conditionalise one kind of thing—things mattering, or, more specifically, me having obligations to my partner—on another kind of thing—the truth or falsity of naturalism.

But many will object that these are not two different kinds of thing, but rather one and the same. “I ought to comfort my partner in distress” entails that “there is at least one non-natural truth” since the former just *is* a non-natural truth. So *of course* any time the latter fails to obtain the former would fail to obtain. But that’s not objectionable; it’s trivial. It’s nothing like saying that my partner doesn’t matter if the Mets don’t win. It’s much more like saying that “my partner doesn’t matter if the person with whom I’m in a relationship doesn’t matter,” or “my partner doesn’t matter if no one in my family matters.”

However, this response begs the question against those who reject non-natural realism, or, as Shamik Dasgupta says, those who advance this objection are not “playing fair” (Dasgupta 2017). Borrowing an argumentative strategy from David Lewis, Dasgupta urges that the non-naturalist is free to posit any *sui generis* non-natural whatnots she likes, but she must play fair in *identifying* these whatnots with goodness (or any other normative property). He challenges the non-naturalist: “Don’t call any alleged feature of reality ‘goodness’ unless you’ve already shown that you have something we should promote” (Dasgupta 2017, p. 301)—or, as I would say, shown that it’s the sort of thing upon which we should conditionalise moral commitment.

The non-naturalist realist seemingly makes substantive claims when he or she posits non-natural moral truths and describes what it takes to discover them. And it isn’t obvious that these seemingly substantive claims apply to first-order moral

claims. By comparison, as Dasgupta points out, water *cannot* be the chemical element Hg (mercury), because water is a clear, colourless liquid and can be detected by seeing if it boils at a 100 °C, and mercury is a silvery and opaque liquid that boils at 356.73 °C. Likewise, even the non-naturalist will admit that parsimony considerations and questions about the explanatory properties of moral facts are at least *in principle* relevant to the question of whether moral naturalism is false. But I've been arguing our norms of moral evidence legislate that these considerations could not in principle be relevant to the question of whether I ought to comfort my ailing partner, or whether anything matters.

Of course, someone who had antecedently accepted that water is Hg would insist that boiling at precisely 356.73 °C *can* be evidence that a substance is water, since it is evidence that a substance is Hg. But that's not playing fair—you can't call your substance "water" unless you have something that is clear and colorless and boils at 100 °C. Similarly, someone who had accepted non-naturalism will be able to insist that parsimony considerations are at least in principle relevant to the question of whether anything matters. Now, as we all know, one philosopher's *modus ponens* is another's *modus tollens*. But I hope my audience will agree that, on the face of it, the claim that *my partner's mattering is not dependent on the existence of some non-causal, judgment-independent state of affairs*, is at least as plausible as the proposal that *"I ought to comfort my partner" states a non-natural fact*.

8 The practical life of nihilists

Another response attempts to counter the accusation of betrayal by questioning the connection between moral beliefs and practical stances. Arguably, coming to believe that my partner "doesn't matter" would only be a betrayal of her if that belief had some practical implications, as it would were it actually to erode my commitments to her.

But maybe that is not the case. After all, ceasing to accept my former moral judgments needn't motivate me to act *contrary* to them. It's entirely conceivable that someone might come to judge that nothing matters, but still be motivated to act in an ostensibly moral way most of the time. Accepting moral nihilism doesn't force someone to behave immorally.⁴

But this response can only take us so far. Even if a moral nihilist continued to go through the motions of the moral life, it's hard to see how her motivational profile would be entirely unchanged. After all, even if I have myriad motivations for behaving decently towards my partner, at least one of those motivations stems from the judgment that *she matters*. Though abandoning that judgment needn't lead me to

⁴ I thank Matt Bedke and Veronique Muñoz-Darde for pointing out to me the possibility of responses along these lines. Bedke draws a parallel with Camus, who famously argued that life could still be worth living, even though it is meaningless. But Camus recognised that denying meaning to life had wide-ranging practical implications. Life *could* still be worth living, but only a very particular kind of life - many otherwise attractive forms of life would indeed be undermined by the denial of meaning (Camus 1955/1942).

treat her contemptuously, it would change my practical stance towards her for the worse. She could rightfully complain that this eroded commitment is a repulsive response to the mere fact that there are no non-natural truths.

Even the most hard-boiled “externalists” about the relation between moral judgment and motivation generally admit that there is at least a strong connection between moral judgment and practical motivation in ordinary, decent, rational agents.⁵ So even if moral nihilism doesn’t force us to become monsters, it surely does make a practical difference. It makes a difference in my relationship with her whether or not I judge that my partner matters. And we should not, morally, conditionalise that practical difference on the truth or falsity of naturalism.

9 Normativity and necessity

I have objected to the conditional premise—that if there were no non-natural normative truths, nothing would matter. But the non-naturalist will object that she denies the antecedent because there are non-natural normative truths. And, furthermore, she claims that this is a *necessary* truth—not a logical necessity, admittedly, but a metaphysical one.⁶

Why worry that some conditional has an unpalatable consequent if its antecedent is necessarily false? Parfit’s conditional looks bad, because it looks rather like the Mets Fan’s conditional. But it is in reality very different, because the antecedent of the Mets Fan’s conditional can be realised! Such conditional propositions are *retrievably true*, but non-nihilistic conclusion follows, and my partner has nothing to fear.

But this misunderstands what we do when we criticise people for the way they conditionalise their moral beliefs. Remember the example of the Kantian and Act-Utilitarian. The Kantian objects to the Utilitarian not because the latter *actually* endorses lying, but because he *would* endorse lying were things different.

This example goes beyond the actual, but it does not take us beyond the realms of possibility. But I submit that we can quite sensibly judge people for their conditional attitudes towards impossible scenarios.

As everybody knows, unicorns are supposed to have certain magical powers. So they are at least *nominally* impossible. But it bespeaks a brutish mentality if you are prepared to torture a captive unicorn. Our attitudes towards *metaphysically* possible scenarios are also morally significant. Laurie Paul has argued that a philosopher’s assessment of the moral status of zombies indicates whether she properly appreciates the moral significance of phenomenal experience—if you think that zombies deserve just the same moral status as normal agents, you undervalue phenomenology.⁷ But philosophical zombies are metaphysically impossible. Even

⁵ Rosati (2016, §3.2) contains an excellent overview of the debate on this point.

⁶ Fine (2002) and Rosen (MS) are exceptions—they argue that normative necessity is distinct from metaphysical necessity. For a criticism of this view, see Lange (2018). For an important argument against the idea that metaphysical modality deserves a privileged or “absolute” status, see Clarke-Doane (forthcoming).

⁷ Cutter (2018) makes a similar claim.

attitudes towards the logically impossible are morally assessable. Doctor Who is not only unreal, he is logically impossible, since he is a Time Lord and one of the defining features of Time Lords is that they can travel in time and change the past. But it's still an indicator of wickedness if someone is prepared to torture the Doctor—and, of course, we do take the Doctor's enemies to be wicked.

More importantly, we should get clear about the modal force of the non-naturalist's claim. She says that certain non-natural normative facts are metaphysically necessary, and I agree that her construal of what non-natural normative facts *are* entails this conclusion—if *there are* any non-natural normative facts. Likewise, it follows from the theist's notion of God that, if there is a God, then God exists necessarily. But in neither case does this admission force me to conclude that God, or non-natural facts, actually exist. Just saying of a claim, that *if* it is true, then it is necessarily true, should do nothing to reassure us that it is, indeed, true.

This is significant, because both the theist and the non-naturalist should concede that it is possible that they are wrong in the sense that they might rationally come to change their minds. Their rejection of naturalism is, or should be, taken as defeasible. After all, many rational people have made the transition, and converted to atheism and naturalism. The non-naturalist and the theist need to plan for the situation where they become naturalists and atheists, just as I plan not to sacrifice Isaac on God's say-so even if I become a theist. I think we should judge these theorists negatively if they were to conclude that they *ought* to become nihilistic and abandon their moral commitments were they to realize this epistemic possibility by coming to accept naturalism.

10 Reasons to act and reasons to change

My arguments bear some resemblance to the famous “wrong reasons” objections made by Williams and other anti-theorists against Utilitarianism (and “moral theory” more generally) (Williams 1973 and elsewhere). According to that objection, utilitarianism requires us to take the wrong considerations to count as our reasons for action. I ought to comfort my distressed partner because she is my partner or perhaps simply just because she's in distress; but according to the Utilitarian, the correct reason is that so doing will maximise utility. According to the Utilitarian, we only really have reason to be partial to our family and friends because we are especially good at making them happy.

Railton (1984) responds to Williams' objection by claiming that utilitarianism *doesn't* require us to take utility considerations as our *reasons* for action—at least not as our motivating reasons. A good utilitarian can still be motivated by the thought that *she is my partner*. Utility considerations are taken into account only inasmuch as I regard it as a *counterfactual condition* on my taking the partnership relation as a reason, that partiality towards partners does in fact maximise utility in general or over the long term. But this thought can be placed to one side in everyday deliberation. As I've argued, this kind of conditionalisation is endemic in moral life.

My objection may look similar—the *fact* that there are non-natural truths is surely not the reason I ought to comfort my partner. The fact that she is my partner

is the reason! But a response analogous to Railton's is available to this objection as well. Non-naturalists have insisted that non-natural truths are not reasons. Scanlon would say, in the case discussed, the *reason* is my partner's distress. What is *non-natural* is the relation of *reason-giving* or *counting in favour of* that stands between her distress and my acting.

However, my objection is *not* analogous to Williams' objection. I don't think that the non-naturalist is identifying the wrong reason *to act*. Rather, my claim is that she is identifying the wrong reason *to change her mind about what counts as a reason*. Scanlon says that the reason for my comforting my partner is a natural fact, but it only *counts* as a reason in virtue of standing in the non-natural reason-giving relation. If there are no non-natural facts, then, according to Scanlon, I would be mistaken to count it as a reason. So discoveries about the existence or not of non-natural facts *are* still being taken as a reason to change one's view about what reasons there are. I don't think commitment to my partner should be even counterfactually conditionalised on the question of naturalism, and the non-naturalist thinks that it should.

11 Robust and relaxed

Some will object that my arguments only tell against *one* kind of non-naturalist realist. "Robust" realists, like Enoch (2011), accept substantive ontological commitments. But many prominent defenders of non-naturalism, such as Nagel (1986), Scanlon (2014), Dworkin (1996) and Parfit (2011), have asserted that their claims carry no positive ontological weight—they are "relaxed" realists. Asserting the existence of a non-natural, judgment-independent moral truth does not, apparently, involve any ontological claims.

This is murky territory. Expressivist quasi-realists have long urged that talk of moral truth can be detached from ontological commitment. And I have a lot of sympathy for expressivism. Pragmatists make similar claims (indeed, expressivists are sometimes called "local" pragmatists). And I have a lot of sympathy for pragmatism. Peace having thus, apparently, broken out, what is there left to object to in non-naturalism?

In truth, "relaxed realism" is rather confusing. The philosophers flying under the banner of "relaxed realism" are insistent that they *reject* the very aspects of expressivism and pragmatism in virtue of which talk of truth without ontology makes sense. And they do not seem to offer anything to take the place of these expressivist and pragmatist theses. So I'm inclined to doubt that relaxed realism is a coherent position at all. But this question is by the way. After all, my objection has nothing to do with the ontological cost of non-naturalism.

However relaxed realists interpret the existential claim "there are non-natural facts," so long as they want to remain recognisably non-naturalist and realist, that existential claim will be the wrong thing upon which to conditionalise moral commitment. Whatever non-natural facts are supposed to be, they must be entirely independent of what anyone happens to think—otherwise positing them is not yet realism. Even if we perfectly follow the norms of investigation implicit in our moral

outlook, the resulting view might still fail to correspond to non-natural reality—after all, our moral outlook might have been false from the start. Even “relaxed” realism makes sense of scepticism. I could be entirely correct about all the relevant natural facts concerning my relationship with my partner and yet quite wrong about whether there is a non-natural relation of reason-giving between her distress and my comforting her.

This is precisely the kind of skepticism that, I claim, makes no sense from within our moral perspective. According to the non-naturalist, if the All Seeing Oracle told us that there are no non-natural moral facts, *whatever* those are, then we ought to drop all our commitment to morality as groundless. But so construed, non-naturalism is a morally objectionable position to adopt. On any available characterisation of the “non-natural facts,” learning they don’t exist could not be a good reason to abandon commitment to our partners.

12 Metaethics versus ethics

My arguments against non-naturalist realism are *normative ethical* arguments. I think we shouldn’t accept a key premise in that view—the conditional premise—because it is morally unacceptable. But, some philosophers will argue, this gets the order of priority the wrong way around. Non-naturalist realism is a metaethical view, and metaethics is a theoretical discipline, prior to, or independent from, normative ethics. We argue for the truth of non-naturalist realism on linguistic or metaphysical grounds—from the fact that moral claims appear to be cognitive, assertoric descriptions of states of affairs or whatever.

However, this descriptive project cannot force us to accept Parfit’s conditional. Even if non-naturalist realism did capture the metaethical presuppositions of ordinary moral talk (which I doubt), there are other metaethical views available. A “revolutionary” metaethics would propose that we *cease* speaking as though we were describing an order of non-natural facts, and start speaking as though we are expressing practical attitudes, or identifying solutions to practical problems and so on. Clearly, such a language could *in principle* be used for the regulation of conduct. Indeed, if familiar moral language and practice does attempt to describe a non-natural reality, then that practice is inconsistent, since it also seems to prohibit us, morally, from conditionalising our commitment to morality on such *outré* facts. The revolutionary’s proposal allows us to preserve our ideas about what counts as a good reason to change our minds in ethics.

This reasoning is, of course, an instance of the is-ought proscription. The fact that we *do* speak one way does not entail that we *should* continue to speak that way. If metaethics has a normative upshot, as I have argued that it does, then no metaethical claim can be substantiated on purely descriptive or theoretical grounds (Väyrynen 2018). The best that can follow from such arguments is an anthropological claim—that some humans happen to speak and think in such-and-such a way—but that is not a claim about how *ethical talk as such* has to be.

One argument that attempts to show we *must* accept a form of non-naturalism appeals to authority. If ethical language “merely” expressed our attitudes, or

pointed out the solutions to practical problems we happen to have, or tracked the goals that animals like us happen to have, then it would lack some special kind of normative authority. But claims about who or what has authority are, of course, themselves normative claims, and criticisable as such.

When a bratty child complains that his teacher or guardian cannot tell him what to do because “you’re not my dad,” he is marking out fatherhood as a basis for authority. The bratty child can (and surely will) be criticised for this because fatherhood is *not* the only legitimate source of authority in this context.

Likewise, if a philosopher insists that expressions of distinctively human and humane emotions, or identifications to the solutions of shared practical problems, *lack authority*, I think she should be criticised. Why *should* we hold out for some other source of authority? I think it denigrates the human, natural, affective world to say its claims lack authority. Why should we care more about what a non-natural order of judgment-independent facts happens to support than we care about what we want, need and feel? To the extent that our morality is humanistic, I think we can reject the argument from authority.

13 Dodging the conditional

David Lewis once suggested that while non-naturalist realism seems to be a default assumption in ethics, if we found out that there were no non-natural facts, we might, rather than succumb to nihilism, adopt naturalism, fictionalism or expressivist quasi-realism as “second best” accounts (Lewis 2005). I suspect that this is an accurate description of the psychology of many non-naturalists. If they discovered that naturalism were actually true, they would not become nihilists. They would simply adopt another metaethical theory. What basis do I have for criticising such philosophers?

This question is based on a false presupposition. My objection was not that non-naturalists *are psychologically so disposed* that they will abandon their moral beliefs if the come to believe in naturalism.⁸ The target of my argument is not the psychology of actual non-naturalists, but rather what is implied by non-naturalism. If I argue that your theory has immoral implications, it’s no defense to respond that you are not entirely convinced in your theory anyway, and would abandon it if the circumstances were manifested in which the theory demanded immoral things. If I argue that utilitarianism is immoral because it would require you to push a fat man off a bridge if there were a runaway trolley traveling underneath a suitably populated span, it’s obviously no defense to say that you are not entirely convinced of utilitarianism and would abandon the theory if you found yourself on the fat man’s bridge. These responses miss the nature of the objection.

I’ve claimed that we *ought not* abandon our moral commitments even if we become convinced (as many of us already are) that naturalism is true. That is a highly plausible first-order moral claim. Not to accept it would be a moral failing.

⁸ This is a major point of divergence from the approach adopted by Bedke (forthcoming).

To judge that the truth of naturalism is no good *reason* to adopt nihilism is something over and above simply being psychologically disposed not to become a nihilist if convinced of the truth of naturalism. I suspect many non-naturalists would actually like to endorse this position. They would like to say that they *should* abandon their metaethics if their metaphysics is false, and become expressivists or naturalists or pragmatists or fictionalists. But they cannot, consistent with their own view, explain what moral grounds they would have for doing so. How can there be a moral basis for accepting morality in a naturalistic world, if moral claims need non-natural truthmakers, and in such a world there are, *ex hypothesi*, none available?

And there is another challenge. If one thought that naturalism or expressivism *were* able to account for the normative authority of ethics, to provide what “really matters”, why bother accepting non-naturalism in the first place? If moral claims could be “true” and authoritative even in a world without non-natural facts, why posit facts of this kind? The conditional-dodging approach robs non-naturalism of much of its motivation.

Indeed, this approach undermines a supposed strength of non-naturalism. According to the non-naturalist, cultures and societies can be radically wrong in their ethics when their views don’t correspond to the non-natural truths. That these alternative ethical systems actually track the extensions of alternative moral concepts, or help their populations solve shared problems, or are just the kinds of principles that it would be useful to invent for those people in their situation, does not, for the non-naturalist, provide a vindication of these systems if they fail to track the non-natural facts. But on what grounds can the non-naturalist then think that *her* moral views could be vindicated by a naturalistic, pragmatist or fictionalist reinterpretation of them? If a failure to correspond with non-natural moral reality falsifies the moral views of alien ethical cultures, *every* positive moral view, however central to our culture, would be falsified by the complete and total absence of non-natural facts.

14 Conclusion

I think my arguments show something about the relationship between ethics and metaethics. After all, nothing I have said does anything to show that there *are* non-natural facts. How could it, when all my arguments have been normative? My claim is that we shouldn’t really *care* whether there are such facts. That the normative question—of whether to conditionalise moral commitment on the existence of the non-natural facts—is antecedent to the existential question of whether there *are* such facts, shows how deep the is-ought divide runs. Any “metaethical” theory that makes claims about what moral talk, thought and practice *must* be like in order to be authoritative is really just a part of normative ethics.

Blackburn once said that finding some argument that would compel even the recalcitrant amoralist to be moral was “the holy grail” of Ethics.⁹ Most philosophers

⁹ Such an argument answers “a wish that the knaves of the world can not only be confined and confounded; but refuted-refuted as well by standards that they have to acknowledge...it is still, tantalizingly, there as a goal or ideal, the Holy Grail of moral philosophy, and many suppose that all right-thinking people must join the pilgrimage to find it. We sentimentalists do not like our good behaviour to

probably accept that, as a psychological fact, no argument could possibly do this. But I think the case is stronger. Even in *rational* terms, no argument can do this. Pointing to the existence of non-natural normative facts wouldn't rationally compel anyone to accept morality unless they had *already* accepted that it matters whether or not there are such facts. So nothing could argue the amoralist—even the *rational* amoralist—into morality. As Humeans, constructivists, sentimentalists, pragmatists and others have long argued, it's only possible to justify morality from *inside* morality. In that sense, realism is in no better position than any other theory. The real question, I think, is what, from the perspective inside morality, would justify *abandoning* morality—becoming nihilistic or amoral. As I have argued, there are some morally acceptable answers to this question. But the answer that non-naturalist realists must offer is not one of them.

Acknowledgements I would like to thank Philip Kitcher, Justin Clarke-Doane, Christopher Peacocke, Simon Blackburn, Sharon Street, Robbie Kubala, Max Barkhausen, Christian Coons, Pekka Väyrynen, Shamik Dasgupta, Matt Bedke, Stephen Findlay, Aaron Zimmerman and audiences at Columbia University, Bowling Green State University, Temple University, University College London, McGill University, the University of Leeds, and the 2018 American Philosophical Association, The University of Edinburgh, Pacific Division for their helpful comments.

References

- Bedke, M. (forthcoming). A dilemma for non-naturalists: Irrationality or immorality? *Philosophical Studies*.
- Benacerraf, P. (1973). Mathematical truth. *Journal of Philosophy*, 60, 661–679.
- Blackburn, S. (1993). Errors and the phenomenology of value. In *Essays in quasi-realism*. Oxford: Oxford University Press.
- Blackburn, S. (1996). Response to Dworkin, BEARS symposium archive, posted 11/11/96. Accessible at: <http://www.brown.edu/Departments/Philosophy/bears/9611blac.html>.
- Blackburn, S. (2010). Must we weep for sentimentalism? In *Practical tortoise raising*. Oxford: Oxford University Press.
- Camus, A. (1955/1942). The Myth of Sisyphus. In *The Myth of Sisyphus and other essays*, tr Justin O'Brien. New York: Alfred A Knopf.
- Clarke-Doane, J. (2012). Morality and mathematics: The evolutionary challenge. *Ethics*, 122, 313–340.
- Clarke-Doane, J. (2016). Debunking and dispensability. In S. Neil & L. Uri (Eds.), *Explanation in ethics and mathematics*. Oxford: Oxford University Press.
- Clarke-Doane, J. (forthcoming). Metaphysical possibility and absolute possibility. *Synthese*
- Coons, C. (2011). How to prove that some acts are wrong (without using substantive moral premises). *Philosophical Studies*, 155, 83–98.
- Cutter, B. (2018). The metaphysical implications of the moral significance of consciousness. *Philosophical Perspectives*, 31, 103–130. <https://doi.org/10.1111/phpe.12092>.
- Dasgupta, S. (2017). Non-naturalism and normative authority. *Proceedings of the Aristotelian Society*, 117(3), 297–319.
- Dworkin, R. (1996). Objectivity and truth: You'd better believe it. *Philosophy and Public Affairs*, 25(2), 87–139.
- Enoch, D. (2010). The epistemological challenge to metanormative realism. *Philosophical Studies*, 148(3), 413–438.

Footnote 9 continued

be hostage to such a search. We don't altogether approve of Holy Grails. We do not see the need for them." (Blackburn 2010, pp. 127–128).

- Enoch, D. (2011). *Taking morality seriously: A defense of robust realism*. Oxford: Oxford University Press.
- Field, H. (1989). *Realism, mathematics, and modality*. Oxford: Basil Blackwell.
- Fine, K. (2002). The varieties of necessity. In T. S. Gendler & J. Hawthorne (Eds.), *Conceivability and Possibility*. New York: Oxford University Press.
- James, W. (1956). The moral philosopher and the moral life. In *The will to believe*. New York: Dover.
- Lange, M. (2018). What would normative necessity be? *The Journal of Philosophy*, 115(4), 169–186.
- Lewis, D. (2005). Quasi-realism is fictionalism. In M. E. Kalderon (Ed.), *Fictionalism in metaphysics*. Oxford: Oxford University Press.
- Nagel, T. (1986). *The view from nowhere*. Oxford: Oxford University Press.
- Nowell-Smith, P. H. (1954). *Ethics*. Harmondsworth: Penguin Books.
- Parfit, D. (2011). *On what matters, Volumes I & II*. Oxford: Oxford University Press.
- Railton, P. (1984). Alienation, consequentialism, and the demands of morality. *Philosophy and Public Affairs*, 13(2), 134–171.
- Rosati, C. S. (2016). Moral motivation. In N. Z. Edward (Ed.), *The stanford encyclopedia of philosophy* (Winter 2016 Edition). <https://plato.stanford.edu/archives/win2016/entries/moral-motivation/>
- Rosen, G. (MS). *What is normative necessity?*
- Scanlon, T. M. (2014). *Being realistic about reasons*. Oxford: Oxford University Press.
- Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies*, 127, 109166.
- Väyrynen, P. (2018). Normative commitments in metanormative theory. In J. Suikkanen & A. Kauppinen (Eds.), *Methodology and moral philosophy. Routledge studies in ethics and moral theory*. Routledge (forthcoming)
- Williams, W. (1973). A critique of utilitarianism. In *Utilitarianism for and against* Cambridge University Press, Cambridge.
- Williams, B. (2006). *Philosophy as a humanistic discipline*. Princeton: Princeton University Press.

Practical Reason, Sympathy and Reactive Attitudes

MAX KHAN HAYWARD
Columbia University

Abstract

This paper has three aims. First, I defend, in its most radical form, Hume’s scepticism about practical reason, as it applies to purely self-regarding matters. It’s not always irrational to discount the future, to be inconstant in one’s preferences, to have incompatible desires, to not pursue the means to one’s ends, or to fail to maximize one’s own good. Second, I explain how our response to the “irrational” agent should be understood as an expression of frustrated sympathy, in Adam Smith’s sense of sympathy, rather than a genuine judgement about Reason. We judge these people because we cannot imaginatively identify with their desires and attitudes, and this is frustrating. Third, compared to the standard cognitive view, my account better explains the nature of our criticism of the “irrational,” and, by portraying “irrationality” as a cause of upset to other people, provides a better *normative* basis for being “rational.”

This paper has three aims. First, I defend, in its most radical form, Hume’s scepticism about practical reason, *as it applies to purely self-regarding matters*. Second, I explain how our critical response to the “irrational” should be understood as a reactive attitude, expressing frustration at being unable to sympathize, in Adam Smith’s sense of sympathy, with the irrational—we simply misidentify this affective response as a cognition about Reason. Third, I argue that this offers a fundamentally interpersonal basis for the value of practical rationality which is more normatively compelling than the standard cognitivist view.

What is self-regarding practical reason? Whilst some of our thought is concerned with representing the way the world is, some of our thinking relates to *what to do*—this is practical, rather than theoretical, thought.¹ Practical thinking deals with motivation, desire and intention.

It’s often thought that, just as there are rational constraints on how we can think, reason and infer about the world, there are rational constraints on practical deliberation, intention and desire. Agents are bound to them *even when no-one else is affected*. For example, the following pick out norms of practical rationality:

- (1) You should not discount costs you will incur just because of where they happen in the future (*no temporal discounting*).
- (2) You should not randomly change your preferences and goals over time (*constancy*).
- (3) You should adopt the necessary means for your ends (*instrumentality*).

- (4) You should not endorse incompatible ends (*consistency*).
- (5) You should desire what you take to be your own greatest good (*maximization*).

I don't deny that these are attractive principles, most of the time. But I want to deny that they are requirements of anything that deserves the title of Reason, and that they have the special binding force often attributed to them. Rather, I think that these norms gain whatever authority they have from the role they play in permitting and facilitating interpersonal relationships. In other words, you don't owe it to Reason to obey the rules, nor do you just owe it full stop, or owe it to yourself. Instead, these norms pick out ways that we need to behave if we want others to be in sympathy with us, in Adam Smith's sense. For Smith, to sympathize with another is to be able to imaginatively identify with her attitudes. Achieving this kind of sympathy, I claim, is necessary in order to maintain social relationships. Thus, rather than the norms of practical reason somehow standing prior to social or ethical norms (and, in the eyes of many, deserving a greater claim to objectivity), they are grounded in interpersonal ethical norms (such as those of benevolence and justice and so on). The normativity of intrapersonal practical rationality is far less obvious, I claim, than that of interpersonal morality. Hence, I try to show why we can find at least some justification for compliance with the norms of practical rationality by appeal to interpersonal considerations, reversing the common view about the priority of self-regarding and other-regarding normativity.

Now in saying that these principles have some force or appeal, but denying that they are *rationaly* obligatory, a certain perplexity arises. What does it mean to say that violating these rules is normally wrong, but not irrational? This is especially pressing in the contemporary philosophical scene, where talk of reasons has become the premier model of normative criticism—rather than being “wrong”, “bad” or even “vicious”, many philosophers think that the fundamental ethical error is to contravene one's “reasons.” At times, talk of reasons threatens to swallow up all evaluative thinking, interpersonal Ethics included. So in what more limited sense of *irrational* is it not irrational to violate 1–5?²

My view, of course, is descended from Hume, who famously wrote:

‘Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an Indian or person wholly unknown to me. ‘Tis as little contrary to reason to prefer even my own acknowledge'd lesser good to my greater, and have a more ardent affection for the former than the latter. A . . . passion must be accompany'd with some false judgment, in order to its being unreasonable; and even then ‘tis not the passion, properly speaking, which is unreasonable, but the judgment. (Hume 2000, BookII, Part 3, Section 3)

In other words, there is no such thing as *practical* reason. But my argument is not Hume's. For when Hume claims that it is not against reason to intend or prefer in these ways, he just means that it is not a *contradiction*, in the sense that two assertoric sentences can be a contradiction. This argument seems too cheaply

won—if you think, as most do, that to desire or intend something is not to have a mental state with descriptive or assertoric content, then *of course* it can't involve this kind of contradiction.

Rather, in looking for supposed facts about practical rationality, I take it that philosophers train their attention on certain possible patterns and structures in the desiderative or intentional states of an agent, and consult their intuitions for a reaction of *wrongness*. This needn't be seen as mysterious—after all, when confronted with the thought of an agent who asserts “Obama is bald and Obama is not bald”, we can immediately see that something has gone wrong with the agent's theoretical thinking. According to many, it is simply a matter of logic that this is wrong, and it takes nothing more than thinking about the structure of the beliefs in the right way to see that this is so. By analogy, the rationalist about practical rationality will say that we can simply tell, by thinking in the right way about the pattern of desires or intentions in the agent who violates the norms 1–5, that something has gone wrong with them.³ (Of course, there are some philosophers for whom even the acceptance of rational knowledge concerning logical laws is problematically mysterious, and so they will not find the analogy to theoretical rationality vindicating for practical rationality.) These immediate responses are *cognitions*, and, in the case of our reactions to cases of supposed practical irrationality, they detect the normative facts. In looking for violations of *self-regarding* rationality, we try to detect wrongness *abstracting away* from the effects the agent's intentions and desires might have on other people.

My view is that the responses we have when we consider cases of agents who violate 1–5 are not, in fact, cognitions detecting normative facts, but rather affective reactive attitudes which philosophers have misidentified. In saying this, I take it that when we experience a reaction to some situation, it is not internally luminous that *this is an intuition*, in the sense of being the delivery of some reliable norm-detection mechanism. All that is internally salient is that we have a reaction which disposes us to make certain kinds of judgements. But to label this an *intuition* requires philosophical work, not mere introspection. Furthermore, along with most sentimentalists and expressivists, I assume that it need not be internally salient, or obvious on immediate presentation to the layman, whether the judgements to which our reactions dispose us are genuine cognitions or descriptions at all. Here I am again in the company of Hume, who insists that “calm passions” can be introspectively “confounded” with the descriptive or representational deliverances of reason, and that it takes philosophical subtlety to distinguish the two. (Hume 2000, Book 2, Part 3, Section 3)⁴ Likewise, it is true that speakers of our language are inclined to utter “that's irrational”—or words to that effect—when they confront violators of 1–5, yet it is open to the philosopher to interpret this utterance not as an assertion, a description of the object's lapse from rationality, but rather as a kind of protest, reproach, or articulation of disappointment. This is my view—the reaction we experience when faced with violators of the supposed norms of practical reason, in the cases where we do feel an urge to criticize them, is fundamentally one of *frustration* or *annoyance*,

and our utterances are expressions of these reactive attitudes, not genuine assertions.

I could just argue directly for that claim, by showing the psychological plausibility and explanatory virtue of this theory. But I also think there is an indirect argument, for there are instances where we consider cases of types 1–5 and do not find a critical reaction in ourselves.

The Rules of Rationality

Temporal discounting. The classic example of this is the agent who chooses larger costs in the far future over small costs in the immediate future—like someone who takes out bad debts or procrastinates. Now of course the Humean can call this irrational if it's based on the self-deceptive belief that the debt will not have to be paid: that would be a form of *theoretical* irrationality—believing against the evidence. And in real cases, it does often seem that it's a cognitive deficiency that underwrites such imprudence—agents convince themselves they'll be wealthier in the future or blind themselves to the inevitability of paying up, in order to shore up their immediate inclinations. But we only detect *practical* irrationality if we feel an intuitive criticism of the agent who *consciously accepts* that she will have to pay up, and still chooses the higher distant cost.

But there is a problem: we need a conception of a cost, such that *my choosing A over B* doesn't entail that A was, for me, more valuable. After all, there's nothing irrational in the ravenous man choosing a loaf of bread today over a warehouse full of the stuff next week.

Derek Parfit asks us to imagine a man with “Future Tuesday Indifference” (call him Indy) who prefers *enormous* amounts of pain on a Tuesday to *tiny* amounts of pain on any other day. (Parfit 2011) But is pain really a stable currency, such that we can insist that Indy will have a greater cost, something *more disvaluable to him*, even though we know that he consciously and clear-headedly chooses it?

Sharon Street suggests that this is unclear. (Street 2009) Allow, for the sake of argument (it's actually empirically questionable)⁵, that unpleasantness is *intrinsic* to the sensation of pain. It's nevertheless true that much of what makes pains so unpleasant is the fear preceding them, the anxiety they provoke, and the way they overwhelm our attention. It appears that various mental manipulations, from specific forms of neurosurgery to advanced powers of meditation, can rob pain of much of this baggage. Imagine that Indy, for complex aetiological reasons, takes this attitude towards his pains *on all and only Tuesdays*. Perhaps he is from a race evolved on a planet that was exposed to inescapable, agonizing but non-fatal cosmic rays 24 earth hours out of each 168, such that developing a disposition to be indifferent towards pain during this period was hugely advantageous. Then it doesn't seem strange or irrational for him to prefer enormous Tuesday pains. They may be more intense, but they will matter less to him.

In that case pain, just like bread in my example above, gains much of its (dis)value from our attitudes towards it. If I prefer one pain to another, that's reason to say

that it was *less disvaluable for me*. Now in cases where our attitudes towards a single thing *change* over time, it's hard to say what its "true" value to me is. In Parfit's original thought experiment—unlike Street's version—Indy only feels indifference to pains that fall on *future* Tuesdays. When Tuesday comes they are just as agonizing as anything. Is Indy's fault that he chooses what he now values, but knows he will in the future disvalue? Surely this is not irrational. The mere fact that I will regret a choice is not a reason not to make that choice. A woman may know that, at the time of labor, she will feel pain and discomfort so great as to make her regret getting pregnant in the first place, but that doesn't mean she acts irrationally in choosing pregnancy—why should the attitude towards labor she has *in the moment of labor* trump the attitude she has at any other time?

Some might object that this would be rational only if we presume that the future regret is merely transitory—our present selves are nevertheless rationally bound to the attitudes of their future selves so long as they are enduring. On such a view, the rationality of a present action might be contingent on foreseeing that, in the future, I will endorse my decision *more* than I regret it. But even this seems wrong. A rich man may know that giving his excess wealth to the Opera will bring him a future of anxiety and regret: without the cushion of a great fortune the benevolence that now motivates him will be shackled and consumed by avarice. But if he determines that this act will bring his life a meaning it would otherwise lack, it doesn't seem irrational. Just as it is common to employ promises and commitments to bind ourselves, like Odysseus at the mast, when we foresee that we will waver from our intended course of action, so it seems reasonable to set myself irrevocably upon a project I value highly, even if I know that the cost of the undertaking is a lifetime of regret. This, perhaps, was the attitude of the band and art group The KLF (also known as the K Foundation) who famously "burned a million quid"—a performance art piece meant to underscore the absurdity of money, which members of KLF regret to this day. Can the prospect of personal regret make the creation of art irrational? I think that it cannot. It doesn't seem that I, now, am under any rational obligation to conform my current values to what I, in the future, will value, endorse or regret.

Perhaps the assertion that the agents in these examples act rationally trades off the fact that their sacrifices are incurred in the service of life projects and ideals. But even this is not necessary. Various commentators on the present paper have pointed out that, really, the only thing unusual about Future Tuesday Indifference is that the willingly incurred misery falls on a *Tuesday*—judging from the headachy and nauseated groans emitting from a million beds across the Western world on an average weekend morning, we might judge that Future *Sunday* Indifference is a fairly common phenomenon.⁶ After all, it would be hard to maintain that the pleasures of drunkenness and well-lubricated social intercourse are really commensurable on any common scale with the pains of a hangover, and even if they are, it's far from obvious that they always or often *outweigh* the misery of the morning after. And it's certainly no secret that ample indulgence in certain forms of liquid refreshment brings about this consequence. Nevertheless, after a week of monkish virtue at my desk, an evening of intoxicated indulgence issues a compelling call, consequences

be damned. Pain and regret are simply the costs of the pleasure I seek, which preference is not, I think, *ipso facto* irrational. I have been known to make the exchange; I have no confidence that the pleasures always outweigh the pains in any sense; and I am not irrational.

Maybe what's wrong with Indy is the *very fact* that he changes his preferences towards Tuesday pain without reason. This suggests that his fault is not really temporal discounting, but **Inconstancy**. But there's nothing wrong with changing your preferences arbitrarily—as Simon Blackburn points out, inconstancy in preferences can be the key to a successful life in various respects—randomly preferring oily fish one day and white fish the next may be a path to health. Unpredictable people can be charming and compelling. (Blackburn 2013) Of course, excessive inconstancy does tend to make an agent *self-thwarting*—plans are laid, toil is invested in putting them in motion, and then the whole project is dropped: after years of study and long shifts, the newly qualified doctor decides she would rather be a poet, and promptly applies for MFA programs. But whilst some people respect the resolution of those who choose a life plan and stick to it with tenacity, I for one find nothing to object to, on rational grounds, with the notion of the butterfly intellect, flitting from project to project, completing nothing. Indeed, Luc Bovens asks us to imagine an agent who values a *soi-disant* “Bohemian” lifestyle, who sees stability and commitment as a trap, the symptom of a bourgeois starchiness, valuing her own will-o'-the-whisps unpredictability as an integral part of her character. (Bovens 1999) It's hard to see what's rationally (as opposed to morally) wrong with this.

To be sure, some people insist that a valuable life must instantiate a certain kind of unified, linear narrative structure, and there is an argument to be made that *this* cannot well be combined with such inconstancy. Galen Strawson calls the former view the *Ethical Narrativity Thesis*. (Strawson 2008) Whilst Strawson acknowledges that this is an overwhelmingly popular view amongst philosophers—he identifies Alisdair Macintyre as a key proponent—he argues, rightly I think, that those who legislate the indispensability of this kind of narrative unity⁷ for all people are inappropriately, perhaps narcissistically, extrapolating from their own psychologies and their own prudential preferences (dare I say prejudices?). Against them, he arrays a constellation of characters who have lived valuable lives despite their disjointed, relationships with themselves: Montaigne, the Earl of Shaftesbury, Sterne, Coleridge, Stendhal, Hazlitt, Ford Madox Ford, Virginia Woolf, Borges, Iris Murdoch, AJ Ayer and Bob Dylan along with Strawson and his parents are, he claims, psychological “Episodics,”⁸ all of whom would sympathize with the sentiment of Henry James, when, reflecting on an early work, he wrote “I think of . . . the masterpiece in question . . . as the work of quite another person than myself . . . a rich . . . relation, say, who . . . suffers me still to claim a shy fourth cousinship.”⁹ Unlike the Strawsons Senior and Junior, I do experience and value a narrative unity to my own life. But I don't have the vanity or intolerance to suppose that this is the only worthwhile way to live. If another person prefers inconstancy to the extent of narrative incoherence, then good for her.

Another form of self-thwarting arises when an agent fails to abide by the norms of means-ends reasoning. Many philosophers, and more economists, have accepted **Instrumentalism**—the view that means-ends reasoning is the sole norm of practical rationality. Indeed, some have even attributed this view to Hume,¹⁰ misled, I think, by his saying that an:

affection can be call'd unreasonable . . . when in exerting any passion in action,
we chuse means insufficient for the design'd end . . .

But of course, Hume assumes that the problem with the agent is that she *believes* that the means are sufficient for her ends, for he continues:

and deceive ourselves in our judgment of causes and effects. (Hume 2000, BookII, Part 3, Section 3)

Now if instrumental claims were nothing more than claims of the form “if you don’t do A, you won’t get B” I would have no problem with them, and nor would Hume, but they also wouldn’t be normative claims of practical reason. For these claims don’t say what we *should* do in any sense—they just state a causal relationship. Claims of practical reason have to be claims about what you should *desire*. Of course, Hume thinks that reason can lead us to correct false beliefs, and so if we are *misguided* about means-ends claims, and *as a result* form desires for things *qua* means, then, in a sense, something irrational has gone on, although he is quite clear that this can only be called a matter of *practical* irrationality in a “figurative and improper way of speaking.” (Hume 2000, BookIII, Part 1, Section 1) since, “Tis not the passion, properly speaking, which is unreasonable, but the judgement.” (Hume 2000, BookII, Part 3, Section 3) Properly speaking, this is a problem of theoretical rationality.

Normally, when we correct our belief, Hume thinks we will also change, as a matter of fact, our desires: “The moment we perceive the falsehood of any supposition, or the insufficiency of any means, our passions yield to our reason without any opposition.” (Hume 2000, BookII, Part 3, Section 3) This is largely correct. But it is just a descriptive psychological claim. To consider whether there is a *norm* of practical reason, we need to consider the case where an agent *knows* her means are not sufficient to her end, but still refuses to change.

I’m unpersuaded that there’s anything irrational here. I have various goals that cannot be achieved except by means that I don’t wish to take up. I want an end to racially-motivated murders by policemen in America within the next 10 years, but I may realize that the only means of doing this by hiring a team of assassins to perform a clandestine cull of the force. More prosaically, I may want a meatball sub, and know that to get it I need to cross the street in the freezing cold. In either case, what is irrational about not intending the means to these goals of mine?

Indeed, it is so perfectly obvious that merely desiring to A doesn’t place you under any rational requirement to desire to B (when B is the necessary means to A) that many philosophers will suspect the critic of instrumental reason of having played some trick. The requirement of means-ends rationality is occluded, they will say, because I have failed to bring in sufficient amounts of heavy-duty

normative language in framing the scenario. It is not merely *desiring* A that places me under a rational requirement to desire to B (when B is the necessary means to A). It is desiring to A *as an end* that generates the requirement to desire to B. Alternatively, the problem is with the choice of attitude—we should switch out “desire” for “intention”—it’s irrational not to *intend* to B if I intend to A (when B is the necessary means to A).

Phrased this way, the normativity of instrumental rationality does seem a little more plausible. But we must be careful to distinguish questions of linguistic usage—definitions or stipulations—from genuine norms of practical rationality. For example, when Kant talks of an end, he says that it is *analytic* that when something is your end you intend the means. My goal here is not to engage in Kant scholarship, but merely to point out that if we assent to something like this, in the *contemporary* sense of “analytic”, then it will preclude the possibility of there being any requirement of practical rationality here. If it’s a matter of *what it means* to desire something as an end that you must desire the means to that thing, then there’s no possibility of criticizing someone for desiring something as an end, and failing to desire the means. *Ex hypothesi*, the target of your criticism didn’t desire the thing *as an end*, and hence the criticism is inapt.

Most contemporary philosophers, I think, don’t use the notion of an “end” in this way, but rather use it to pick out something desired in itself—my “end” is the object of what Parfit calls a “telic desire.” This weaker sense of “end” doesn’t make it a mere matter of definition that I desire the means if I desire the end. But neither does it support any normative requirement that I *ought* to desire the means if I desire the end. Indeed, as Dewey¹¹ pointed out in his writings on the “reciprocity of means and ends,” it is simply bizarre to conduct our practical thinking by first picking out a list of ends, and then just mechanically latching onto the means to our ends in order to furnish ourselves with a full set of desires. Ends are not neatly delineated objects picked out by the narrow headlights of an unswerving will. When we see the ideal of practical thought as the operations of an agile intelligent, not a rigid machine, it is clear that “ends” can quite reasonably stand to be fluidly reinterpreted and endlessly reevaluated in the light of their context—and in particular in the light of what needs to be done to get them.

In a similar vein, the concept of intention wavers between being either so robust that it builds instrumentally-structured attitudes into intentions as a pure matter of definition, or too weak to support a norm of instrumental rationality. Some might claim that I don’t count as genuinely *intending* an end unless I *intend* the means. But, again, if this is a stipulation about the meaning of “intend” then it’s not a norm of practical rationality—no-one could be criticized for failing to intend the means to their intended end, since, *ex hypothesi*, they would not be intending the end.¹² To avoid this consequence, we might take a more permissive definition of intention. Perhaps we will then say that to count as intending some end I must simply intend to take *some* path towards it, beyond mere wistful wishing. We may ask—what if I foresee that the path I choose will not, on this occasion, take me where I aim to go? Is it irrational if I then proceed? Not at all. People can knowingly, and indeed quite admirably, set out on doomed

quests, valuing honorable striving more than underhanded success. The path I choose for bringing about racial justice is rational persuasion and peaceful protest. Even if I foresee that this will not be enough, it does not seem irrational for me nevertheless to continue—I am merely doing my best in the sad state of the world.

Some will say that while it's not true that we're obliged to adopt the means for every end of ours, it's still irrational to continue *both* desiring the goal *and* refusing to adopt the means. I don't have to do what's necessary to reach any goal, but if I won't do what's necessary, I need to drop the goal. Perhaps, then, the problem is one of **having incompatible desires**—simultaneously desiring or intending two things that can't both be had? But what's wrong with desiring immediate racial justice, and refusing to perform the necessary murders to achieve it? Mutually incompatible goals are part of the richness of life—I want to be ethical, and cheerful, even though I know these things may conflict. I can remain committed to the wellbeing of each of my sons, even when they are at each other's throats. Single-mindedly streamlining our intentions in the name of “rationality” is nothing short of bizarre to me.

Now, having incompatible desires is still a case of self-thwarting—you won't do as well, in terms of getting a greater proportion of your desires satisfied, as a more streamlined agent. We might see all forms of self-thwarting as instantiations of the generic problem of **failing to maximize your own good**. Since violations of 1–4 are often cases of self-thwarting, we might then try to elevate maximization into an ur-rule of rationality, from which all the others can be derived, or at least partially justified.

But what is “my good”? We've already seen that sensations like pain get much of their subjective importance from our attitudes towards them. It is very strange to say that pain is bad for you even in cases where you do not mind it at all. So we may go with the orthodox economist's view and say that our good is defined by our preferences. In this way, we may say, against Hume, that Reason *does* require us to adopt our own acknowledged good. But what *are* preferences? These same economists endorse the principle of revealed preferences—a preference is a disposition to action, so I prefer whatever I choose in a fully factually-informed situation.

But, as Blackburn points out, the combination of these two views makes it impossible for economists to do what they advertise as their competence—that is, dispense normative advice about what would be rational to do in matters of practical deliberation. For, if I *chose* a self-thwarting arrangement of desires, doesn't it just follow that this was my preference, and that I preferred inconstancy, or commitment to each of two incompatible goals, more than I cared about *getting as high a percentage of my preferences satisfied as possible*? If you interpret a fully informed agent as having gone against her preferences, it follows that you misinterpreted what her preferences were! (Likewise, in the classic prisoner's dilemma scenario we are told that it's rational to play Defect—but that's only because we assume that our costs are only measured by years in prison. If we measure the agent's costs in terms of her preferences, then if she plays Cooperate, that's to say, refuses to turn in her partner, then we have to interpret her as valuing Cooperative behavior more

than freedom—and why is this irrational?—in which case she’s not really playing the game “Prisoner’s Dilemma” as modelled by economists, and hence the advice doesn’t apply). (Blackburn 2013)

Practical Reason then comes to nothing more than the injunction “do what you most want to do.” *Contra* the economists, I think that we ought to accommodate the possibility that an agent fails to do even this. We might plausibly think of desires and preferences as mental states of the agent which are typed by their *typical* propensity to bring about certain kinds of action, or at least attempted action. This leaves it open that desires may anomalously fail to bring about their typical products, even in the absence of other, over-riding desires. Extrinsic causes may make even a powerful occurrent desire anomalously ineffectual at moving an agent, just as a strong muscle may in sudden spasm drop its habitual load. I may want nothing more than to flee from the oncoming bear, but find myself rooted to the spot; I may desire overwhelmingly to tell my companion how special tonight has been, yet be distracted by a sudden explosion from the kitchen; faced with a split-second decision, my natural disgust at pushing fat men off bridges may be so far foremost in mind that, by the time my guiding passion for Utilitarianism muscles itself forward through the medley of desires, the moment—and the trolley—have passed. In such cases, we needn’t infer that my strongest desires are not precisely those I have claimed.

Does inching away from the theory of revealed preference in this manner allow us to resurrect some substantive norms of maximization, from which to derive the further norms of practical rationality? Hardly. It is only as a matter of anomaly that we can thus think of desires as failing to bring about their proprietary effects; if the failure is frequent then we should interpret the agent as having another, countervailing desire, or simply not desiring strongly enough after all. But if the norms of practical rationality are supposed to be norms that we can hold agents to—which they should, but might not, intend to honor, and by reference to which we can criticize them when they lapse—then irrationality cannot be a matter of mere anomaly. You would not berate the hiker for failing to flee from peril, nor the lover for springing from the table, romance forgotten, nor the flustered utilitarian, finally confronting the situation she has so long considered, who pauses for one fatal beat too many. If our desires or intentions are anomalously inefficacious, then the problem is not a matter of what we desire or intend. The upset occurs causally downstream from our actual desires or intentions. The agent does not intend for her desires to be causally inefficacious, and there is no point criticizing her if they are, because there is nothing, after all, that she could have done about it.

The confusion, I think, arises from the legacy of using the term “*akrasia*” to refer to two quite different things. Sometimes, “*akrasia*” is supposed to refer to the failure to desire that which you think best; this is at least a candidate for being a norm of rationality in the sense under consideration, since the problem is a question of *what* the agent desires or intend, and is something for which it could make sense to criticize her. But other times, “*akrasia*” is used to refer to failures to do what you desire or intend. One interpretation of this corresponds to the notion of anomalous failure mentioned above. Avoiding *akrasia* in this sense may be desirable, but it

could not be a norm of practical rationality. There is no change in the agent's intentions or desires that could prevent it; it is a failure that lies beyond reasonable criticism, beyond the bounds of agency.

We might interpret the injunction to maximize as requiring agents to *get* the sort of preferences of which the largest possible percentage can be fulfilled—which might entail abandoning any preferences or scruples which we care about more than we care about *the fact of getting a large proportion of our preferences fulfilled*. An advantage of this view is that injunctions against inconstancy, counter-instrumentality, incompatible preferences and future-indifference can be derived from it. But it is a very odd theory—it says that the value of my life is measured by the proportion of preferences of mine that get satisfied *even if that's not what I care about*, and tells us to *change* our preferences to start wanting things we don't otherwise care about. It's hard to see what could motivate someone to endorse this principle, but it is at least clear how to comply with it—brainwash, drug or otherwise condition yourself so that you care about nothing much apart from breathing, eating, drinking, sleeping and defecating. These preferences are delightfully compatible, and there will be no temptations to counter-instrumentality or inconstancy. Your chances of scoring high in terms of proportion-of-preferences-satisfied will be very good indeed, and so, in this sense, you will be “maximizing”. I doubt many philosophers would defend such a position, but it is important not to confuse this view, which does imply an injunction against self-thwarting preferences, with the more minimal economist's version of maximization, which doesn't. But we can set it aside now, because obedience to this norm is incompatible with living a worthwhile life. If this were what practical reason enjoins, I would want nothing to do with it, and nor, I imagine, would you.

Another interpretation of the norm of maximization moves even further from the theory of revealed preferences. Some philosophers distinguish between an agent's mere desires and preferences on the one hand, and her genuine values on the other. Perhaps rationality requires that we maximize, not the satisfaction of our desires, but the fulfilment of our *values*—the agent's values, not her mere preferences, set the yardstick for her good. I ought to do what I most value; failure here corresponds to the other notion of *akrasia* mentioned above. Since my values don't always manifest themselves in effective desires, this injunction is not so empty as the requirement to maximize preference-satisfaction. We can interpret an agent as desiring, intending and acting against her values, and criticize her on that score as irrational.¹³

There are many different accounts of the distinction between an agents' values and her mere desires, and lack of space precludes a full discussion here. Values might be desires that the agent desires to have, or that are especially stable, or that withstand reflective scrutiny (or other things besides). Each of these accounts answers to some aspect of our pre-theoretical notion of valuing; and yet they are mutually incompatible—they will often attribute different lists of values to a given agent. But even where these conceptions agree on what an agent's values are, we can think of cases where individuals appear to go against what they value, but we feel no criticism of them, and indeed may think they are making their lives go better. Spontaneity often means acting against one's settled and reflective values, and the

spontaneous decision does not always reflect a shift in higher-order desires. And yet it seems wrong to me to criticize all such spontaneous decisions as *irrational*. Indeed, we think a degree of such spontaneity can be good for the agent. Live a little! we say, don't overthink it—do something you'll regret! Always subordinating one's desires, passions and preferences to the authority of serious or settled values may signal a monastic rigidity; obsessive, even cold. The occasional unruly holiday from the strictures of our mores can be a much-needed antidote to the monotony of self-control.

Indeed, occasionally acting contrary to my deeper or settled values may provide me with vital experiential evidence needed to reform those same values. When the conservative Mormon, Joe Pitt, in Tony Kushner's *Angels in America* gives in to the temptation of his lust and follows a gay man into Central Park to declare his desire, he acts in a way that, from his deepest evaluative perspective, is despicable and shameful. Every settled, reflective, internally endorsed value in Joe decries his homosexuality as a failing and perversion. And yet, driven onwards by his *akratic* passion, he discovers a love and tenderness in another man's arms that move him to reject his previous precepts. Had he acted always on his values, he would never have experienced and understood the fulfillments which they prohibited to him, and on the basis of which he comes to change his outlook. In letting his passions override his values, Joe lacked a certain kind of self-control; and yet I do not think we would criticize this as irrational, and it hardly seems contrary to his own good.

But perhaps most importantly, this norm of value-maximization could not ground the other norms. When my values clash with one another, there is often no unique maximizing solution. That is what is shown by the examples I discuss. When I refuse to assassinate the police to bring about immediate racial justice, it is my values that clash, not just desires or intentions. The parent of warring sons values the well-being of each, and so has incompatible values. The norm of value-maximization does not force me to be instrumental here, because instrumentality just as much as counter-instrumentality means sacrificing the realization of one of my values on the strength of the other. Likewise, the norm of value-maximization says nothing that would make the rich man defer to his future self. If his current values tell him that his life will have a meaning it would not otherwise have if he donates to the opera, why should he defer to his foreseen future regret—even if, perhaps, his future self no longer values opera and greatly values the excess wealth with which he is considering parting? Here, current values clash with future values. If you have clashing values, then there is no unique way to maximize their fulfilment. If you value, either synchronically or diachronically, things that are incompatible, and if your utility function is a matter of satisfying your values, then you *have* no one utility function. The future-discounting, inconstant, counter-instrumental, or incompatible route is just as good, from the perspective of maximization, as the “rational” solution. Of course, not everyone has values that would thus justify them in being counter-instrumental or in temporally discounting. But it would hardly be a resurrection of the norms of practical rationality to say “observe the instrumental norm, unless you have some value that would lead you not to do so.”

As with preferences, we could promote maximization by changing our values. Should the parent of warring sons *change* what she values, abandoning the commitment to one of her sons, so that maximization picks out a unique outcome? Should the rich man simply cease valuing opera to lessen the tension between his current and future values? Again, this is absurd. Having clashing values is just one impediment to maximization. Often, the impediment stems from an individual value itself, because the goals we have set ourselves are too high, our aspirations too lofty, our commitments too demanding. We could have easily-maximizable values just by dint of valuing what is easiest to have, by setting our sights lower. But most of us value *retaining* our current values far more than we value the prospect of having *other* values that would be easy to fulfil. Both aiming low and abandoning conflicting values are ways of setting yourself up for a less challenging life. It seems extremely implausible that a less challenging life is always better for the liver of that life than a more challenging one.

The fact that our values are not fungible—cannot just be swapped out, revised or streamlined to make it easier to satisfy them all—is in a way tragic, but that fact is just part of what it is to be a creature that cares about things. The idea that norms of maximization constitute any deep insight into how to live well is to miss the fact that to value is to be open to disappointment, to commit to a project is to make real the possibility of failure. If the good life for an individual person is wrapped up in what she values and commits to, it is simply shallow philosophy to say that she should change her values to make her life easier and thereby better. The change itself would be a loss, an abandonment. Few of us have a single utility function, and we could not get one without excising much of what makes us who we are. Philosophers have a tendency to fetishize coherence, but it is far from obvious why this should be a guiding light of practical life. Occasional incoherence is so deeply a part of the human condition, arising so naturally and persisting so frequently, that the demand for coherence requires serious justification. I can't see what purely *self-regarding* reason we have, or even could have, to make such a transformation. My view is that the rich man, the bohemian, the doomed quester, the mother of warring sons—all these may well be living lives as good as possible for them.

Some people will say that you should obey the rules of practical rationality because you have an obligation to yourself to live a Good Life, and that what makes a life good isn't a matter of your preferences, or what you value, but is instead an objective fact (from an Objective List). In other words, there are weighty substantive prudential goals that we're all required to pursue that are totally independent of our desires and values, and all the rules of practical reason are derivative from them. We'll come back to this, but I want to turn now to my positive view.

The Psychology of Reacting to the “Irrational”

We have seen that, in certain cases, we feel no urge to criticize the agent who violates the norms of practical rationality. The Bohemian, the rich man who gives his money to the opera, I when I refuse to assassinate the police in order to bring about racial justice—none of these agents are doing anything *wrong*. Of course, such

characters might still be rather baffling—without knowing about Indy’s strange past, for example, it would be difficult to understand his motivations; without understanding that the Bohemian finds constancy to be unacceptably starchy and bourgeois, she might seem flighty to the point of insanity. And in many cases an onlooker might feel rather sad that the values of the agent led her to such a conflicted or self-thwarting situation. I shall discuss later how it is possible to feel sad for the “irrational” agent without in any sense criticizing her, without feeling that she has done anything wrong or ought have different intentions or desires. But I take it that this is likely to remain an unusual case.

For, despite all I’ve said, we may still find it hard to shake the feeling that there is generally something *wrong* with people who violate norms like 1–5—people who prefer torture on Tuesday to pinpricks tomorrow, who take out loans for medical school only to pursue poetry, who pick out their dream job but can’t be bothered to post the application, who want both to have a perfect body and to eat Kraft Mac’n’Cheese daily, who always play Cooperate in Prisoner’s Dilemma situations, leaving themselves open to endless exploitation. The urge to criticize such people is almost inescapable.

My view is that we should acknowledge this response, but we do not have to take it at face value. That’s to say, we don’t have to assume that our response to violators of 1–5 constitutes a *belief* that these agents have betrayed the standards of practical reason.

Rather, if we consider the conception of sympathy offered by Adam Smith, we can see our judgement of the “irrational” as the expression of a reactive attitude. According to Smith, sympathy involves imaginative identification with the situation and circumstances of others. I imagine myself in your situation, and see what attitudes I would have were I you—in contemporary terminology, I perform an “off-line simulation” (off-line because I won’t necessarily act on the attitudes summoned up in my breast when I imagine myself to be in your position). When our brother is on the rack:

By the imagination we place ourselves in his situation, we conceive ourselves enduring all the same torments, we enter as it were into his body, and become in some measure the same person with him, and thence form some idea of his sensations, and even feeling something, though weaker in degree, not altogether unlike them. (Smith 2009 Part 1, Section 1, Chapter 1)

Smith’s notion of simulation mainly focuses on placing myself in your *external* situation, but, as we’ll see, the account gains strength when we acknowledge the ability of thinkers to simulate one another’s *internal* situation. But, and this is the distinctive claim of Smith’s moral psychology, we (normally) have an enormous *desire* to observe correspondence between the real attitudes of others and the attitudes we imaginatively find in attempting to identify with them:

Nothing pleases us more than to observe in other men a fellow-feeling with all the emotions of our own breast, nor are we ever so much shocked as by the appearance of the contrary. (Smith 2009, Part 1, Section 1, Chapter 2)

Most of the time, agents who violate the norms 1–5 are very difficult to sympathize with. When I imagine having spent years of my life and tens of thousands of dollars on medical school, it is hard to imagine *not* wanting to be a doctor and *wanting* to be a poet. Now if Smith is right—and I think he is—and we do normally want to observe correspondence between our simulating selves and the realities of others, then this desire will be frustrated when confronted with people who temporally discount, or don't pursue the means to their ends, and so on. This frustrated attempt to sympathize is a cause of irritation and upset—instead of the pleasure of correspondence, the “irrational” agent offers the onlooker the aggravation of discord.

This irritation would be amplified if we add, to Smith's desire for correspondence, benevolent desires on behalf of others such as Hume supposed to be within us all. For whilst I may not be able to imagine having the preferences of the “irrational” agent, I can still wish that her desires—in as far as I discern them—go fulfilled. I form a partial view of her wellbeing, and long for that. But, since “irrational” agents are self-thwarting, my benevolent desire will itself be thwarted. The thwarting of our well-meaning desires for others is a further source of frustration. Hume changed his views on the extent of our benevolence, only in the second *Enquiry* asserting the existence of a generalized, albeit oftentimes extremely weak, desire for the wellbeing even of strangers.¹⁴ According to that view, all things being equal (for example, where it costs us nothing either way), we tend to prefer that others be benefitted, even when they are totally unknown to us. But my account is neutral as to the existence of such a universal sentiment. Our benevolence to others varies in degree, from the intense to the extremely weak—or even nonexistent—and, as we shall see, this variation explains a variation in our responses. Where benevolence is powerful, it amplifies the annoyance we feel at those who thwart themselves.

Thus, the immediate reaction of criticism that we gain when faced with violators of the norms listed at the start needn't be seen as *detections of facts about the norms of rationality*, but as *expressions* of a reactive attitude caused by frustrated sympathy. And, in fact, we can now see *direct* evidence for the Humean/Smithian theory. That theory predicts that we don't just criticize “irrational” agents, but feel annoyed by or upset by them. And we *do* feel this way when we imagine buffoons who undergo distant agonies rather than present trifling inconveniences. But it is very unclear *why* we should have this *affective* response if irrational agents were merely summoning up a cognitive detection of a normative fact. Why get angry at them if they're merely wronging themselves or betraying reason?

Furthermore, our response to the “irrational” agent admits of variability, even when the *fact* that she has violated 1–5 remains fixed, suggesting that our response doesn't just track the fact of norm-violation as such.

First, there is a marked tendency to find less irrational those patterns we exhibit in ourselves. People who change their goals often, for example, seem to have less powerful intuitions about the irrationality of inconstancy than those who single-mindedly pursue one project. Those who pursue diverse projects at the same time seem to find the adoption of incompatible goals more obviously tolerable. And, likewise, the irrationality of counter-instrumentality seems less obvious when I present

you with an end you would never abandon connected to a means you would never adopt. That these judgements should so vary with our own desiderative dispositions is strong evidence that our judgements of “irrationality” are not the deliverances of some normative-fact-detecting faculty, but are rather, as the Smithian suggests, born in the breakdown of sympathy. Of course, most of us know this tacitly—it is an inexperienced drunkard who looks to a teetotaler for sympathy in the depths of his hangover. Unless the clean-liver is a character of unusually expanded sympathies, from that quarter the drunkard had better expect little more than scorn and shame. By contrast, there is a certain communion among the imprudent.

That’s not to say, of course, that the feelings we find when we imaginatively place ourselves in some situation always track those feelings that we have had, or would have, in really occupying that situation—far from it. Even explicitly knowing that one would feel a certain way does not guarantee that one finds the very same feeling in imaginative simulation. Sometimes imagination does not keep pace with reality, even lived reality. This is why it is possible to criticize others for failings that we ourselves exhibit—even to criticize ourselves. In the depths of a hangover it is hard to imaginatively think myself into the situation of accepting just one more drink—even when this is what happened mere hours before, and I know that well. In penitent rectitude I may criticize myself, or others in my predicament. Nevertheless, it’s still true that we are far less prone to find ourselves and other like us to be irrational than those whose desiderative and intentional patterns are entirely alien to us.

Secondly, on being given more information about the lives of the agents, we become inclined to withdraw our criticisms—*even though* these are still cases of inconsistency or inconstancy or whatever. When you gain more biographical (biological!) information about Indy and his strange planet, as you try and think yourself into the lifestyle of the self-consciously inconstant Bohemian, the sense that these people are *wrong* starts to fade. After all, as any novelist will tell you, it is easier to think myself into an alien mindset when I am equipped with facts about the other’s interiority and history. Lily Bart, protagonist of Edith Wharton’s *The House of Mirth*, is a nice case in point. She ruins her own life, wrecking every prospect she has of social and financial redemption after her initial fall from grace, which results in her own destitution and, eventually, death. All this is done in the service of scruples that she can barely articulate. The brilliance of the novel is that, although her desire to remain honorable and independent is strong—which is why she rejects the options offered to her—her love of wealth, leisure and status is also powerful; it would be hard to say that her scruples in any way *outweigh* her more material interests, and thus, her decisions do not maximize her own values, goals or well-being. And yet, although she causes her own destruction through a series of knowing choices, the reader is not drawn to the conclusion that she is irrational, criticizable, foolish. The glimpse we gain into Lily’s interiority is so vivid that we sympathize with her attitudes and decisions, and so through understanding we dissolve our own frustration at her self-harm.

Thirdly, we should on the other hand recognize (as, I think, Smith did recognize—Hume is less clear on this point) that sympathy (in the sense of

imaginative identification) and benevolence are distinct phenomena, and can in some cases pull against each other. Because of this the intensity of our reaction against the “irrational” agent is moderated by two kinds of distance. The kind of distance I considered above is that of imaginative sympathy—it measures the degree to which we are in fact able to understand the other from the inside, and see how we might adopt her attitudes in her condition. But the other kind of distance is a more blindly affective one, and it measures not our degree of understanding but simply the intensity of our desires—desires to find correspondence, and benevolent desires for whatever we take to be the other’s good. I said just before that when we get closer to others, in the sense of sympathetically understanding them better, as we do when their interiors are exposed to us by the novelist’s pen, our reactive response—the reaction that rationalist philosophers had misidentified as an intuition of “irrationality”—gets weaker. As we draw nearer in understanding to Lily Bart, we lose the urge to judge her. But we can be close to another in the *second* sense, without necessarily having an understanding from within. This, I think, is often the feeling of parents towards their adolescent and adult children—in their benevolence towards their offspring they yearn for correspondence and the satisfaction of what they take to be their progeny’s best interests, but they often fall short in comprehending the choices and attitudes of the younger generation. When we are close in this purely affective sense, but without having a full sympathetic understanding, then our judgement of the other grows not weaker but stronger—the sense that this agent is culpably irrational for abandoning medical school for poetry is all the stronger when it is my daughter, rather than a stranger, who is thus inconstant in her projects.

At the other extreme, both affective and imaginative responses are sometimes so etiolated that we feel no reactive attitudes at all. The would-be murderer who will not avail himself of the necessary means to his intended end arouses little ire or judgement in us for his failings of means-ends “rationality”—though we surely cannot understand his motivations, and thus lack Smithian sympathy, we barely care to engage with him in the first place, and so feel little anger or sense of blame. If his counter-instrumentality is seen as irrational at all, it arouses no sanction. With the exception of those philosophers who are rigidly determined to uphold the laws or practical reason, few of us will be able to say, with any real conviction, that he *ought* to have done otherwise. Even in their case, I doubt that the judgement of irrationality is a response to an immediate reaction to the situation, a sense that something has gone wrong and ought to have been done differently; judgements in such outré cases are rather reached by extending the normative strictures we accept elsewhere. We do not gain further evidence that counter-instrumentality is bad because we see that bungled murders are bad; rather, the only motivation to think of bungled murders as bad is a commitment to the claim that counter-instrumentality is bad.

It is compatible, of course, with the standard cognitivist theory that our anger at the prudentially irrational agent should grow as our benevolence grows. Even if it is simply a fact, out there in the world, that this or that describes the good of the other, it is quite possible that I should *care* about the good of some people more than

others. But what the cognitivist will struggle to explain is the way in which one kind of interpersonal closeness amplifies our negative reaction against the inconstant or counter-instrumental agent, whilst another kind of closeness diminishes it. And it is obscure, for the cognitivist, why we should feel almost no judgement at all towards the counter-instrumental murderer. The sentimentalist, by contrast, appealing to the interlocking effects of both Smithian sympathy and Humean benevolence, can offer us a strikingly seamless explanation of the confusing landscape of our reactive judgements.

These three considerations, then—the diminution of our negative reactions when the “rules” are broken in ways that we are ourselves inclined to break them, or by people whom we have come to understand biographically and psychologically, and the amplification of those same reactions when we feel greater benevolence towards the errant subject—point strongly to the sentimentalist interpretation: that our so-called “intuitions” about what is practically rational are really reactive attitudes of frustrated sympathy and benevolence—they’ve just been misidentified by rationalists.

The Normativity of Rationality

Finally, I think that the Humean/Smithian view is a better *normative* basis for endorsing 1–5 than simply claiming they are requirements or obligations of rationality. After all, as I’ve pointed out, it’s pretty rare for someone to willfully violate 1–5—normally cases of temporal discounting or counter-instrumentality are symptoms of a cognitive deficiency—wishful thinking or selective blindness. But if someone *genuinely* doesn’t want to be “rational”, why should the appeal to the requirements of reason move her? Even if rationality did legislate 1–5 as exceptionless imperatives, as I have argued that it does not, this fact could hardly play any role in persuading people to be consistent, or constant, or instrumental. If we tell the Bohemian that she is being irrational in her inconstancy, she may well agree, but go on to insist that she really does not care about our rules of rationality in any case (indeed, she may rather delight in flouting such rules), and urge us to leave her alone, for, after all, she is not harming anyone else. So long as *her* mind does not bridle at inconstancy, what magisterial weight should the invocation of reason carry? Or we may insist that it is constitutive of her being an agent that she obey the rules. But this is either false, or trivial. The bohemian isn’t obeying the rules, and yet she does weigh options and make decisions—it’s just that many of these decisions get overturned. In the minimal sense of “agent” in which agency is required for the mind to direct the body *at all*, she surely *is an agent*. So agency, *in that sense*, doesn’t imply obedience to the rules. Alternatively, we may mean that it is constitutive of *rational* agency that she obey the rules. But if she didn’t care about the rules in the first place, why should she care about achieving that form of agency which is defined simply in terms of obedience to the rules?

This may seem like a “merely” psychological point, and rationalists will object that I am committing the classically Humean error of confusing claims about what sorts of argument might actually persuade real people in the world with questions

of what is objectively obligatory. Even if the inconstant isn't *moved* by the appeal to reason, the rationalist insists, she *ought* to be. But even as a purely normative claim, the invocation of rational obligations, commitments or requirements in the purely intrapersonal case is suspect. In the everyday world, obligations, commitments and requirements fundamentally exist in the space between agents. I make a commitment to you, you require something of me, we compact an obligation one to the other. How can an agent simply owe something *simpliciter*? When the Bohemian objects to her critics "What's it to you? I choose this, knowingly and in full understanding of the consequences, and I am harming no-one else," she is making a *normative* argument to which I see no rejoinder. Until we can produce some real person *to whom she owes it* to be constant, it seems simply like a priggish rule-worship to insist that she must so be. Or shall we say that she owes it, not to another real person, but to Reason? That is very strange. As William James pointed out:

If we must talk impersonally, to be sure we can say that "the universe" requires, exacts, or makes obligatory such or such an action . . . But it is better not to talk about the universe in this personified way, unless we believe in a universal or divine consciousness which actually exists. (James 1956, section II)

James recognizes that impersonal modes of talking about normativity, "There is an obligation" or "It is required" are, in modern atheistic parlance, merely the residuum of a conceptual world in which there always was an actual divine someone to whom things really were owed, "some supreme authority to which individual intelligence was absolutely in bonds," (Dewey 1948) who really did require us to act one way or another. Invocation of Reason as the holder of our obligation to be constant or instrumental, is, and ought only to be, impotent in persuading us to be obey 1–5. Until some real person is affected, the Bohemian may be as inconstant as her passions demand.

Now, many will respond that there is a real person who is affected—the Bohemian herself. They may suppose that agents owe it *to themselves* to respect the norms of practical reason. This picks up the suggestion left off above—perhaps the normativity of 1–5 is not intrinsic to the rules themselves, but stems from the fact that obeying 1–5 helps an agent to achieve the good life for herself, robustly construed, and she is under an obligation to herself to live as well as possibly.

The details of our conception of the good life will matter here. After all, as argued above, if living well means living the kind of life I want to lead, then the person who wants to ignore 1–5 will be living well if she does so. But even with a more concrete conception of the good life, it's dubious that obedience to 1–5 will be sufficiently closely connected to personal success for the latter to ground an obligation to comply with the former. Future indifference can prevent the fear of age and death from casting a long shadow over the rest of life, adopting incompatible goals allows us to participate in a richer and more various diversity of projects, inconstancy can be exciting and counter-instrumentality can be scrupulous. Freedom from fear, engaging in a diversity of projects, excitement—these are all very plausibly ingredients of the life that is objectively good for the liver, if we believe in such a thing, and yet they are got by "irrationality", not rationality.

That's not to say that practical reason and self-interest always part company, and if we must regulate our lives in accordance with some set of exceptionless rules, then the rules of practical reason will probably do better than any other. But, of course, we don't have to pick rigid rules to live by. So this observation won't give us grounds for criticizing the agent who, due to whim or unusual circumstances, chooses to flout the rules on any particular occasion. It is no defense of the traditional theory of practical rationality to say that we should be (say) instrumental "most of the time", or "so long as the agent wouldn't be better off by not being instrumental".

Still, perhaps there is some robust conception of the good life from which obedience to the rules never comes unstuck. And perhaps this is the right image of the life well-led. To be sure, if we make such an appeal to substantive, meaty prudential goals it may be hard to see how we are talking about requirements of *reason* in any limited sense (in any sense where talk of reason doesn't, as mentioned above, swallow up all of normativity)—but at least we will have grounds to argue that obedience to the rules is a *self-regarding* obligation.

However, considered properly, it should be clear that the notion of an obligation owed to myself cannot do the necessary work. The whole point of being the holder of an obligation is that one has a power to release the other—so if I owe something to myself, I can also release myself from that obligation. That's not to say that the language of self-directed obligations is entirely senseless: many philosophers see our obligations to *other* people as extremely extensive and demanding, and so counterbalancing these with a basket of obligations to the self is one way of granting agents with what James called a "moral holiday", without having to suppose that the moral holidaymaker is simply ignoring or flouting his obligations. But even if we recognize obligations to the self for this reason,¹⁵ a self-directed obligation would not be something by reference to which we could *criticize* a violator of 1–5. The Bohemian, if she ever did have an obligation to herself to be constant, must surely be understood to have exercised her right as the obligation-holder, and released herself. If, by contrast, it turns out that I can't release myself from a self-directed obligation, then this is a case where the obligation is not really owed *to* me, as a debt or a promise is owed to me; rather, it is just another case of an obligation which specifies my self in the *actions* required, but is really owed to Reason, with all the strangeness that entails.

We may say that I owe an obligation, not to my current self, but to my *future* self—who, in the case of someone like Indy, will surely not be inclined to release me. But even despite the metaphysical murkiness of this, it strikes me as very important that we *don't* view ourselves in this way—the case of the wealthy man shows how important it is *in being* an agent to do something that a future part of myself would wish I hadn't done, and would even reverse if he could. To view my future self *as mine*, is to view him as someone whose feelings I am uniquely entitled to disregard.

Now, if I *don't* regard my future self as mine, then the sense that I am entitled to treat him as I wish does, it is true, fade. For example, we might tell Indy's story in a different way, so that his indifference is not to Tuesday pain but to him Tuesday *self*—we don't interpret him as mindfully and deliberately deciding that Tuesday pain matters to him less than other flavors of pain, as making a judgement

about what he wants for his own life, but rather as choosing it because he literally doesn't care what happens to Tuesday Indy. But in that case, Indy's choice feels more like a substantive *moral* violation than a prudential one. What is so objectionable about Indy is not that he has miscalculated the judgement about his own self-interest—for he doesn't take the affairs of Future-Tuesday-Indy to be even *relevant* to his self-interest—but that he is prepared to treat *someone* so callously. It's indifference to *people*, not miscalculations of prudence or self-regarding rationality, that we criticize in him. If we endorse such a criticism, it's not a claim about self-regarding practical reason, but about when morality starts—it starts at the breakdown of self and other, and that rupture of identity which only sympathy can bridge.

But the obvious answer to the normative question—the question of what should move us to be “rational” if we're not otherwise so inclined—is already present here, and it brings us back to the Humean/Smithian picture. If there is any obligation to obey the rules of “practical reason” it would have to be an obligation owed *to someone*. It's certainly odd to imagine it owed to some distinct agent, also called Max Hayward. But why not see it as owed to other, real, concrete people then? As we've seen, “irrationality” is something that causes distress to those around us, since it frustrates their benevolence towards us. The argument that you should be rational so as to avoid causing unnecessary upset to others strikes me as far more normatively compelling than the argument that you should do it because it's simply a requirement of rationality, and harder to wiggle out of (by absolving yourself) than the argument that you owe it to yourself. And it has the potential to be persuasive. So long as we care about others, as Hume thinks we generally do, then we'll feel disinclined to snub them by thwarting ourselves. Likewise, the incoherence of “irrationality” causes agents to fall out of Smithian sympathetic correspondence with one another, damaging their ability to enter into and sustain a whole range of valuable interpersonal relationships. A degree of coherence is a necessity for anyone who lives in society with others. Again, this seems like a powerful normative argument for conforming to the norms of practical rationality—if we live in society and maintain relationships then it seems that we ought to do what's necessary to remain relatable and preserve those relationships. Anyone wanting this kind of social life will feel the tug to conform to the rules of rationality, out of broadly moral, or at least interpersonal, concern.

Of course, an obligation to be rational thus grounded will not be absolute. This is hardly surprising—surely only the most hardboiled rationalist would think that the requirement to be practically rational cannot be over-ridden, even by substantive interpersonal moral considerations. My view is that the requirement to be practically rational is itself an interpersonal moral consideration; so it should be clear just why and how it might be over-ridden by more serious concerns. And the obligation to be practically rational will have no basis in agents who live entirely separated from others who might care about them—Robinson Crusoe has no obligation to be practically rational if he doesn't feel like it, at least until Friday turns up. Someone who abandoned society entirely, escaping all social ties and free of all relationships, would no longer have any other-directed reason to remain in

compliance with “rationality”. Of course, those raised in social contexts will have already internalized the perspectives and criticisms of others, so will continue to feel the tug to be practically rational even if they find themselves in a state of isolation where there are no others to be upset, no relationships to damage; but there is at least no harm in this. There’s no reason for Crusoe *not* be rational, if he has internalized the urge to do so. And the account predicts that those who are improperly socialized, who do not successfully internalize the perspectives of others, will not be moved by the arguments that they ought to be practically rational when they don’t otherwise feel like it.

This explains why psychopaths, although deeply self-interested, are also famously imprudent. Along with amoral lack of concern for those around them, they display:

poor behavioral controls and tend to commit crimes from a young age. They are impulsive, irresponsible, and . . . unable to set or stick to realistic goals for themselves or to consider the possible consequences of their actions, which can lead to self-destructive behavior. (Bollard 2013 pp238–59)

The sentimentalist theory is, I think, uniquely equipped to explain why psychopathy leads so unflinchingly to disregard for *both* morality and prudence. Psychopaths do not care about the feelings of others, nor do they want meaningful relationships, and so the sympathetic, benevolent, pro-social motivations to comply with the norms of both morality and “practical rationality” are simply absent in them.

That we can see the normativity of self-regarding practical reason as ultimately grounded in pro-social sympathetic concern should remind us how appealing sentimentalism can be. Rather than the crude instrumentalism, or even moral skepticism, that is sometimes wrongly attributed to them, both Hume and Smith make it clear that normative thought *in general* is something that can only occur in an agent who has internalized the viewpoints of other people. There should be no worry if normativity doesn’t seem to “show up” from a purely objective perspective, for the normative point of view just is the intersubjective point of view:

That we owe a duty to ourselves is confessed even in the most vulgar system of morals; and it must be of consequence to examine that duty, in order to see whether it bears any affinity to that which we owe to society. It is probable that the approbation attending the observance of both is of a similar nature, and arises from similar principles, whatever appellation we may give to either of these excellencies. (Hume 1998 Appendix IV)

Hume’s suggestion is entirely in keeping with the position I have put forth: far from standing before and apart from our moral concern for others, the sense of having a *duty* to ourselves is as much derived from sympathetic interpersonal concern as the most altruistic attitudes. My account, then, helps us to see how sensible is sentimentalism’s insistence that normative thought is ultimately *social* thought—even when thinking about the intrapersonal.

It also invites a spirit of open-mindedness. For we can now see the norms of rationality as something of a negotiation. In the typical case, it seems easiest for the anomalous, “irrational” agent, mindful of the concerns of others, to contort herself

back into line with what is commonly relatable. After all, the others are the majority, and if, as I have suggested, it is normally extremely difficult to sympathize with the “irrational”, then it is asking a lot of the rest of society to require that they make this effort. But it always remains possible that, instead, the onlookers should simply make a better effort of trying to think themselves into the situation of the person they find objectionable. It’s common, I think, to dismiss the preferences and intentions of others as simply irrational—from the disability activists who prefer blindness to treatment, to working class people who squander hard earned dollars on scratch cards—or, if these things aren’t irrational, they are failures of self-regarding obligation. But if we think that there are no norms of self-regard, but only of other-regard, that criticism of another is always an expression of myself, then the option remains to us to try harder to think ourselves into their points of view. It’s true that blindness hinders your ability to achieve many of your goals, and to maximize your preferences generally—but you may value these very limitations. Much as I would like to give *ability* to others, I should be open-minded to the possibility of valuing *disability*—instrumental rationality be damned. And it may be true that playing scratch cards doesn’t maximize your expected utility (and sadly, many people who play them do not understand this fact, and would behave differently if they did). But perhaps you value the mere possibility, the hope of genuine financial security over the certainty of an only-slightly-less-grinding poverty, and if you do this knowingly, I am not sure I should insist that you reform. In these cases, it is not, I think, asking so much of society to make the imaginative stretch of sympathizing with the anomalous agents, whereas it seems to me like a great imposition to require the blind not to value their disability or the poor to abandon their improbable hopes. If I am right, then our ability to see practical rationality in others is limited only by our imagination.

This theory illuminates a further possibility. I mentioned before, parenthetically, that it is possible to abandon the critical stance altogether when we confront people behaving “irrationally”. The very wise, especially when they are engaging with those they truly understand, often feel merely sad at the reality of self-thwarting behavior. In these cases, the irritation, frustration and upset I mentioned above are absent. They have managed to do what is normally so difficult to do—to enter into Smithian sympathy with the “irrational”. Why, then, are they sad? It is sad when people have self-thwarting preferences for exactly the same reason that it is always sad when someone (who we care about) wants what she cannot get. The only difference between the general case of wanting what you cannot get, and the more specific case of self-thwarting, is that the obstacle in the former case is external, and in the latter case internal, to the agent herself. But in neither case is insisting that the agent change her preferences a solution to the problem. If someone has incompatible desires, then she simply will not get everything she wants. This fact is not changed if she later changes her desires so as to erase the incompatibility. If I want large-scale wealth redistribution, I also will not get what I want, and neither does this situation cease to be sad if disappointment and cynicism lead me to stop wanting redistribution. Having desires or values, either synchronically or diachronically, that, as it happens, *cannot* both be fulfilled,

is just like having desires or values that *will not* be fulfilled—both are ways in which agents fall out of step with the world. Rationality gives us no basis to prefer the cost of getting themselves into step to the costs of remaining out of step. Insistence that the self-thwarting agent contort her attitudes into the mold constituted by the norms of “practical rationality” occludes the fact that there is no loss-free solution to her predicament; so long as we have the capacity to stretch our sympathy to encompass the “irrational”, it makes more sense to be sad than to be judgmental.

Finally, it shows that Sidgwick’s fear that nothing could bridge the normative requirements of self-regarding Prudence and other-regarding Ethics is groundless. The Prisoners’ Dilemma is sometimes seen as demonstrating the existence of a conflict between what is practically rational for the individual players, and what would be best for the aggregate, or collectively rational. But now we can see that this is a mistake. There is no rational requirement for players in the Prisoner’s Dilemma to prefer any particular course of action—and so there is no Prisoner’s Dilemma. We see a conflict because, when we sympathetically and benevolently imagine each player individually, we would have each choose whichever path will get him the fewest jail years, and so project onto him a requirement to play Defect, but when we imagine *both* sympathetically and benevolently, we want the fewest jail years for the two, and so feel that they must play Cooperate. But this is not conflict of individual and collective rationality, or of Prudence and Ethics. For, I have argued, there’s no such thing as purely self-regarding practical normativity. The Prisoner’s Dilemma simply illustrates the potential conflict between partial sympathies (towards one player over the other), and impartial ones (aimed at both at once). And that is a very different kind of problem.

Notes

¹ Some people, including me, worry that there is no strict distinction here, but I lay that aside for now.

² The best-known proponents of this approach are Derek Parfit (2011) and Tim Scanlon (2014), but the emphasis on reasons-talk is not limited to those who officially endorse “reasons fundamentalism”.

³ I think many philosophers understand intuitions in something like this way. But my argument does not hinge on any particular positive account of intuitions. Whatever intuitions are, I show that our negative or critical responses do not in fact track violations of 1–5, since there are cases of such violations where we do not find the response. Thus these reactions cannot be reliable detections of any normative facts inherent in violations of 1–5.

⁴ Peirce (1868) makes a similar point.

⁵ See especially Grahek (2007).

⁶ The two commentators who independently suggested this point to me, whose names I withhold to shield them from the scrutiny of puritans, are, perhaps not unsurprisingly, both Humeans. Indeed, it appears that Hume’s circle was one in which the pleasures of the bottle were well-appreciated, so it is not too great a stretch to speculate that he was familiar with this phenomenon.

⁷ Obviously, picaresque novels have their own kind of narrative structure, but “narrativity” as Strawson intends it refers more narrowly to a particular diachronically unified linear structure, one not broken up into discrete and disjointed episodes.

⁸ I quote Strawson's list not to endorse the claim that, as a biographical fact, each person on the list experienced life in a "non-narrative" fashion, but rather in order to illustrate the conception of non-narrativity that Strawson puts forward.

⁹ Strawson's quotation is drawn from Henry James (1864–1915/1999) 1915 pp 562–3.

¹⁰ Millgram, (1995) makes a fuller case for the claim that Hume rejects instrumentalism.

¹¹ See especially Dewey (1981) Vol. 13, pp 210–19.

¹² Finlay (2009) makes a similar point.

¹³ I am grateful to an anonymous reviewer for urging me to explore this possibility.

¹⁴ In the Treatise he claims that "In general, it may be affirmed, that there is no such passion in human minds, as the love of mankind, merely as such" (Hume 2000 Book 2, Part 3, Section 3) but in the second Enquiry he changes his position, saying "No man is absolutely indifferent to the happiness and misery of others. The first has a natural tendency to give pleasure; the second, pain" (Hume 1994, Section V, Part II) and "There seems here a necessity for confessing that the happiness and misery of others are not spectacles entirely indifferent to us." (Hume 1994 Section VI Part I)

¹⁵ I would like to thank Mary Grace Albanese, Max Barkhausen, Simon Blackburn, Justin Clarke-Doane, Stephen Finlay, Philip Kitcher, Robbie Kubala, Christopher Peacocke and two anonymous reviewers for their helpful comments on drafts of this paper. I would also like to thank audiences in Oxford, Cambridge, Rotterdam, Milan, Rome, Cardiff and Sheffield for their helpful discussions, and especially Lorenzo Greco, Tito Magri and James Lenman, who got it twice.

References

- Blackburn, Simon (2010) "Practical tortoise raising" in *Practical Tortoise Raising and Other Philosophical Essays* (Oxford University Press, 2013).
- Bollard, Mara (2013) "Psychopathy, Autism and Questions of Moral Agency" in A. Perry & A. Yankowski (Eds.), *Ethics and Neurodiversity* (Cambridge Scholars Press).
- Bovens, Luc (1999) "The two faces of akratics Anonymous" in *Analysis* 59 (4):230–236.
- Dewey, John (1948) "Some Historical Factors" in *Reconstruction in Philosophy* (Boston MA, Beacon).
- Dewey, John (1981) *The Later Works, 1925–1953*, J. A. Boydston (ed.), (Carbondale: Southern Illinois University Press).
- Finlay, Stephen (2009) "Against All Reason?" in Charles R. Pigden (ed.), *Hume on Motivation and Virtue* (Palgrave Macmillan).
- Grahek, Nikola (2007) *Feeling Pain and Being in Pain* (Cambridge Massachusetts: The MIT Press).
- Hume, David (2000) *A Treatise of Human Nature* (Oxford: Oxford University Press).
- Hume, David (1994) *An Enquiry Concerning the Principles of Morals* L.A Selby-Bigge ed, revised P.H Nidditch (Oxford: Clarendon Press).
- James, Henry (1864–1915/1999) *Henry James: a Life in Letters*, ed. Philip Horne (London: Penguin).
- James, William (1956) "The Moral Philosopher and the Moral Life" in *The Will To Believe* (New York: Dover).
- Millgram, Elijah (1995) "Was Hume a Humean?" in *Hume Studies Volume XXI, Number 1 (April, 1995)* 75–94.
- Parfit, Derek (2011) *On What Matters* (Oxford University Press).
- Peirce, Charles Sanders (1868) "Questions Concerning Certain Faculties Claimed for Man" in *Journal of Speculative Philosophy (1868)* 2, 103–114.
- Scanlon, T.M. (2014) *Being Realistic About Reasons* (Oxford University Press).
- Smith, Adam (2009) *The Theory of Moral Sentiments* (New York: Prometheus Books).
- Strawson, Galen (2008) "Against Narrativity" in *Real Materialism and Other Essays* (Oxford: Clarendon Press).
- Street, Sharon (2009) "In defense of future Tuesday indifference: Ideally coherent eccentrics and the contingency of what matters" in *Philosophical Issues* 19 (1):273–298 (2009).