Max Hayward
Columbia University

Practical Reason, Sympathy and Reactive Attitudes:

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 maximise 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 sympathise, 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.

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

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 (*maximisation*)

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 sympathise 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 *rationally* 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.

[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."

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

[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.

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 criticise 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 cases 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

[4] Peirce (1868) makes a similar point.

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, agonising 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.

---

[5] See especially Grahek (2007).

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 agonising 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 labour, 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 labour she has *in the moment of labour* 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

---

[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.

preferring oily fish one day and white fish the next may be a path to health. Unpredictable people can be charming and compelling. (Blackburn 2010) 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 unity[7] 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

[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.

"Episodics,"[8] all of whom would sympathise 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

---

[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 pp562-3

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

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 realise 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 criticising 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 Dewey[11] 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

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

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 criticised for failing to intend the means to their intended end, since, *ex hypothesi*, they would not be intending the end. 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 honourable 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 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 maximise your own good**. Since violations of 1-4 are often cases of self-thwarting, we might then try to elevate maximisation 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 behaviour 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 2010)

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 maximisation, 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 honour, and by reference to which we can criticise 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 criticising 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 intends, and is something for which it could make sense to criticise her. But other times, "*akrasia*" is used to refer to failures to do what you desire. 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 maximise 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 "maximising." 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 maximisation, 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 maximisation 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 maximise, 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 maximise preference-satisfaction. We can interpret an agent as desiring, intending and acting against her values, and criticise her on that score as irrational.[12]

[12] I am grateful to an anonymous reviewer for urging me to explore this possibility.

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 criticise 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 *aktratic* 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 criticise this as irrational, and it hardly seems contrary to his own good.

But perhaps most importantly, this norm of value-maximisation could not ground the other norms. When my values clash with one another, there is often no unique maximising 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-maximisation does not force me to be instrumental here, because instrumentality just as much as counter-instrumentality means sacrificing the realisation of one of my values on the strength of the other. Likewise, the norm of value-maximisation 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 maximise 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 maximisation, 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 maximisation by changing our values. Should the parent of warring sons *change* what she values, abandoning the commitment to one of her sons, so that maximisation 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 maximisation. 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-maximisable 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 maximisation 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 fetishise 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 criticise 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 criticising 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 criticise 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 sympathise 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 sympathise 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 generalised, albeit oftentimes extremely weak, desire for the wellbeing even of strangers.[13] 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

[13] 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)

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 criticise "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 teetotaller 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 criticise others for failings that we ourselves exhibit - even to criticise 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 criticise 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. Lilly Bart, protagonist of Edith Warton'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 honourable 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 maximise 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, criticisable, foolish. The glimpse we gain into Lilly's interiority is so vivid that we sympathise 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 recognise (as, I think, Smith did recognise - 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 wilfully 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)

James recognises that impersonal modes of taking 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 criticising the agent who, due to whim or unusual circumstances, chooses to flout the rules on any particular occasion. It is no defence 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 recognise obligations to the self for this reason,[14] a self-directed obligation would not be something by reference to which we could *criticise* 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

[14] Indeed, I think we should not, since this approach is really just a way of paying lip service to the idea that agents are governed by their obligations, whilst really releasing us from our obligations to others. It would really be more honest, if we really want to carve out space for self-interest, to simply claim bald-headedly that moral obligations aren't everything.

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 her 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 his Tuesday *self* - we don't interpret him as mindfully and deliberately deciding that Tuesday pain matters to him less than other flavours 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 criticise 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 internalised 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 internalised the urge to do so. And the account predicts that those who are improperly socialised, who do not successfully internalise 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 unfailingly 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 internalised 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 sympathise 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 maximise 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 maximise 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 sympathising 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 behaviour. 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 mould 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 judgemental.

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.

Bibliography

Blackburn, Simon (2010) "Practical tortoise raising" in *Practical Tortoise Raising and Other Philosophical Essays* (Oxford University Press)

Bollard, Mara "Psychopathy, Autism and Questions of Moral Agency" in A. Perry & A. Yankowski (Ed.), *Ethics and Neurodiversity* (Cambridge Scholars Press, 2013)

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)

Grahek, Nikola *Feeling Pain and Being in Pain* (Cambridge Massachusetts: The MIT Press, 2007)

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)

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)*