

Ethics as a Humanistic Inquiry

Max Hayward

Submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017

ABSTRACT

Ethics as a Humanistic Inquiry

Max Hayward

This dissertation argues that ethics is fundamentally mind-dependent. Ethics is invented by humans, to solve the problems that mutually sympathetic agents find in living together. Ethical discovery is the discovery of solutions to the kinds of problems that humans find themselves to face. Views of this kind are familiar, but I attempt to re-orient the debate. Many philosophers see questions about the foundations of ethics as fundamentally theoretical, arguing for one view or another on metaphysical or linguistic grounds. I argue that the question of which metaethical view we adopt is a substantive, first-order moral question. And, contrary to many, I think that first-order considerations speak in favour of a variety of anti-realism. We should reject the search for non-natural, mind-independent, objective moral truths as morally objectionable: it denigrates interpersonal concern, making the significance of moral and practical life dependent upon abstractions remote from what we care about and ought to care about. By contrast, seeing norms of morality and practical rationality as collectively created by processes of interpersonal sympathy shows why they matter, and explains the goals and methods of moral inquiry.

Contents

Acknowledgements.....	i
Introduction.....	1
Bibliography.....	12
1. Practical Reason, Sympathy and Reactive Attitudes.....	15
Bibliography.....	52
2. Nonnaturalist moral realism and the limits of rational reflection.....	54
Bibliography.....	76
3. Immoral Realism.....	79
Bibliography.....	106
4. Sentimentalist Pragmatism Defended.....	109
Bibliography.....	138
Conclusion.....	142
Bibliography.....	148

Acknowledgments

I have been incredibly fortunate to be supported by a wonderful dissertation committee. My sponsor and mentor, Philip Kitcher, has been a constant source of help and inspiration, and a guide in both my intellectual and personal life. He taught me not only to be a better philosopher, but, I hope, a better person. Simon Blackburn has been involved in my development as a philosopher since before I even started at university. His influence runs deep in my thought, work, and interests. Christopher Peacocke has been an excellent advisor, both in my writing and my professional development, and the differences our philosophical views have made our conversations all the more interesting and useful. I was extremely fortunate to have Justin Clarke-Doane join the Columbia Department as I was embarking on my dissertation, and discussions with him have been invaluable in the development of my work. And I am grateful to Sharon Street for her agreeing to join the examining committee, and for our conversations over the years.

Robbie Kubala has been my best friend and intellectual companion since we both came to Columbia in 2011. Almost everything I have thought or written in that time reflects conversations with him. I cannot exaggerate how helpful it was for me to have such a friend. Max Barkhausen has been a great friend and thoughtful critic, and he has always been generous in providing detailed comments on drafts of my work. And Mary Grace Albanese has not only talked through every idea and argument in what follows, but has also read every page. She has not only helped me gain a richer perspective in my understanding of philosophy, but also to develop my own style as a writer.

My family have been a constant source of support and encouragement. I know they always believed that I would eventually gain the PhD and become the first “doctor” in the family, but I know how pleased they are to finally see it completed. I am grateful for their unfailing belief in me.

The Columbia Philosophy Department has been a wonderful place to work for the past 6 years. As well as my committee, the faculty here have always worked hard to make this a welcoming and stimulating environment. The environment here has equally been shaped by the incredibly team of staff in the department. Stacey Quartaro has run a brilliant operation in my time here, and I am hugely grateful to her for making so many things possible in my time at Columbia. I am also grateful to Maia Bernstein and Asha Panduranga, and to Clayton Rains for his invaluable help in the final stages. The other graduate students at Columbia have made this an exciting, warm and welcoming place to work, and it has been a privilege to live and work among them.

The wider philosophical community have created a wonderful environment for me to develop as a young scholar. I owe enormous thanks to the generous comments and feedback I've received from other philosophers at talks, conferences and workshops.

During my time writing this dissertation, I received funding from the Columbia Philosophy Department, Columbia University GSAS, the Benjamin Wollstein Scholarship, the Heyman Center for the Humanities, the New York Council for the Humanities, and the Mellon Foundation. Many thanks to all of these for supporting my work

Introduction

Many ethicists worry that morality is *merely* an expression of our attitudes, emotions, or sympathy for one another, that our values reflect perspectives that are *just us*, that the principles we endorse are *merely* contingent, that the natural world is *all there is*, that ethical inquiry is *no more than* invention, and that the reforms we make in our norms are *mere* change. These worries take place against the backdrop of a supposedly loftier aspiration - that our values, purged of messy human contingency, could, through a process of discovery, aspire to correspond with objective, mind-independent, rationally obligatory, universal normative facts.

My primary target in these essays is not the claim these aspirations are obscure, metaphysically implausible and psychologically unrealistic (although I do think all those things). Rather, I want to deny that they constitute an *aspiration* at all. Or rather, these are things that we *should not* aspire to, because there is no reason to think that they are better than the “mere” alternatives. In the background of these hopes for objectivity, mind-independent truth, rationality and all the rest there is an implicit evaluation - a denigration of the human perspective, of emotion and sentiment, of contingency and of invention, and a concomitant elevation of the “perspective of the universe”, objectivity, rationality, timeless universality and discovery of what is “out there” independent of us. And this is what I object to. I love the human perspective, I celebrate humane emotions, I am glad of contingency and variety, I hold invention as worthy as discovery. If that is all there is, then it strikes me as outrageous to yearn for more.

What I see, then, at the heart of these ethical and metaethical debates, is a clash of values. Blackburn (borrowing from Nietzsche) once said that rationalist opposition to the Sentimentalist

moral theories that descend from Hume and Smith stemmed from a privileging of the “Apollonian” over the “Dionysian” in human life. (Blackburn 2001 Ch4). Sentimentalists do not neglect the Apollonian, but they also have a hearty respect for the human, messy, emotive Dionysian world. A similar opposition is found in classical Pragmatism, in Dewey’s objection to the “intellectualist” elevation of static representational knowledge over practical, problem-solving, world-engaged practical thought. Both these traditions have a core that I think of as *humanistic* - a rejection of philosophy’s long preoccupation with rationality and knowledge of necessary truths in favour of the human practical perspective. And they offer mutual support. Sentimentalism can benefit from the more detailed theory of moral inquiry offered by the pragmatic focus on inventive problem-solving; Pragmatism can benefit from the rich moral psychological theories of the sentimentalists.

But my goal in these essays is to clear the ground for a sentimentalist pragmatism in ethics, not to articulate my own version in detail. My arguments are chiefly focused on the clash of values that I described above, and I aim to defend the humanistic account of moral authority and moral inquiry and expose the hollowness I see in currently popular forms of moral realism and rationalism. As such, my arguments are chiefly negative or defensive, and my constructive proposals are largely suggestive and schematic.

The first paper, “Practical Reason, Sympathy and Reactive Attitudes,” seeks to overturn a near-universal understanding of the role of rationality in practical life. Certain rational norms are supposed to hold momentous authority in regulating practical life. Agents are bound to them even when performing actions that have implications for none other than themselves - Robinson Crusoe may be free from all human laws and interpersonal obligations, but he still owes fealty to the rules of practical rationality. And these norms gain their authority from nothing more than the fact that they are the dictates of Reason. For many philosophers, in response to the question “why should I do that,” the response “because that is the rational thing to do,” is considered the ultimate normative

conversation stopper. A norm enjoining us to obey reason is, supposedly, more obvious and compelling than any other norm.

I deny this. The putative norms of practical reason may constitute perfectly useful advice for a variety of situations, and very often they indeed do have authority over agents. But they are not good advice because they state exceptionless normative rules, and they are not authoritative - when they *are* authoritative - because they find their source in rationality. In fact, to the extent that the norms of practical rationality can be articulated in a form that is neither empty nor hideously gerrymandered, they are norms that an agent living alone might quite sensibly flout, and in doing so deserve no criticism from us. Criticism of the “irrational,” when it is deserved, gains its normative authority not from Reason, but from the practical role that obedience to these norms plays in facilitating human social life.

As I argue, the best descriptive way to make sense of our patterns of response to the putatively irrational, to understand both the cases where we are outraged and those where we tolerate the “irrational,” is to see our criticism as stemming not from reason but from sentiment. We do not perceive that the “irrational” have violated Reason and condemn them for that; rather, in many cases, the “irrational” are simply hard to sympathise with, and thus to relate to - inasmuch as we are embedded in relationships with them, we will be hurt or annoyed by “irrational” agents for undermining these relationships. And this is what our criticism expresses. Sometimes we can so sympathise, and these are precisely the times when we withhold criticism.

These are simply descriptive, psychological claims. But my ultimate purpose in this paper is a normative one. Once we have seen that the norms of rationality are not exceptionless, that sometimes obeying reason does not immediately give us anything of value, we may wonder why we should care about the rules of rationality at all. This is a normative question, and the answer lies in

the account of sympathetic relationships I just gave. We should be rational, even if there is no other reason so to be, in order to facilitate the sympathy that undergirds our relationships with others around us. This, of course, makes the authority of rationality highly contingent. If our capacities for sympathy were different, it would make sense to impose different practical norms; indeed, those capacities are flexible enough that our practical norms - at least at the periphery - can and should remain a site of constant negotiation.

As such, my goal here is to perform a complete reversal of the standing account. Too many philosophers have thought that grounding the norms of practical life and morality in rationality would give a final answer to all questions of authority - that if we could show that reasons requires it, all questions of the form "but why should I do that?" would be answered forever. I aim to have shown the complete opposite. The real question is "why should I be rational?". Sometimes the answer is that you need not be. But where you should, the ultimate normative authority comes from the role of these norms in human, interpersonal life. Obedience to the norms of rationality provides a solution to one of the ongoing problems of human social life - the problem of maintaining the sympathetic ties that make society possible. That the norms perform this practical function is the best possible explanation or normative grounding for their authority that I can think of. Of course, the precise shape of the practical problems we face is determined by contingent facts about humans psychology. But the norms are no less authoritative for the fact that they are contingent. In fact, because they attend to the peculiarities of the human condition, they are all the more compelling.

So rationality does not undergird the authority of practical thought. The second paper, "Nonnaturalist moral realism and the limits of rational reflection," retains my critical focus on the role of rationality in ethics, this time focusing on the theory of moral inquiry. According to many, moral investigation aims to discover purely normative truths, and it is rationality, or, more specifically, rational reflection, that enables it to do this.

My argument is that the investigatory methods that fly under the banner of rational reflection could not be competent to lead inquirers to mind-independent normative truths. In the paper, I mainly focus on the “non-naturalist” conception of mind-independent normative truths, according to which the normative truths are causally inefficacious and ontologically independent of the natural truths. But I suspect that my arguments can be extended to encompass other forms of moral realism that are structurally similar to non-naturalism.

The two main contemporary parties to this debate in moral epistemology are the “debunkers,” who think that naturalistic genealogical considerations on their own are sufficient to rule out the possibility that moral inquiry could lead us to non-natural truth, and the realist-rationalists, who think that rational reflection is enough to counteract any distorting effects of social or cultural evolution, and to guide us to the moral truth. However, there is not enough discussion between the two parties on the question of what, precisely, rational reflection is, and how it is expected to work. According to the debunkers, evolution shows that our current beliefs are probably so far from the non-natural truth - if there is such a thing - that reflection could not possibly provide a corrective.

I am broadly on the side of the debunkers, but I deny that they have given the right argument against the non-naturalist’s epistemology. The genealogy of morality does not, as they claim, show that our epistemological situation is utterly hopeless. Rather, two things are true. Naturalistic considerations rule out the possibility of any quasi-perceptual source of moral knowledge. And genealogical considerations show that, even if some of our moral beliefs are true, in the non-natural sense, many others are surely false - and that includes our moral intuitions and assessments of what is plausible. That is to say, while we may start with some true beliefs, neither perception, intuition or assessments of plausibility can help us sort the true from the false.

This is why it is absolutely imperative to consider in detail what rational reflection is supposed to be, and that is my major contribution in this paper. If reflection simply sought to remove contradictions from our moral belief-sets, then we could see why it would be truth-conducive - many philosophers agree that contradictions cannot be true. But, as I show, literal contradiction in normative judgements is far less prevalent than is commonly thought, and these contradictions are easily resolved without making major revisions to our ethical viewpoints. It is more plausible to see rational reflection as seeking to promote *coherence* relations of mutual support that go beyond mere non-contradiction. But this is a problem for realists, because there is no reason to think that the moral truth needs to be highly coherent in this sense. In fact, it is quite reasonable to think that many legitimate values pull against one another: so “true” morality might well be full of messy conflicts.

One justification for coherence-seeking in epistemology appeals to the role of the theoretical virtues. On this view, we do not seek coherence directly as an end in itself, but because it promotes the virtue of our theories. But then the question reasserts itself - why seek virtuous theories? Realists will have to argue that virtuous theories are more likely to correspond to the mind-independent normative truths. But they can provide no reason, in ethics, to think that this is so. Again, goes the retort, why rule out the possibility that the moral truth is messy, conflict laden, even sometimes tragic? The realist cannot explain to us why coherent, theoretically virtuous moral outlooks are any better than those they supersede - which latter, they must admit, are at least partially erroneous. So the realist seems committed to a kind of pessimistic anti-theory.

But my goal is not to deny that coherent, theoretically virtuous ethical outlooks are often superior; rather, my point is that the realist cannot explain *why* they are. This is the point in my argument which, I think, extends to most realist theories, not just non-naturalism - although I do not argue this in detail in the paper. Rather, to explain why we ought to prefer coherent, simple theories which minimise conflict, it is better to appeal to a pragmatic theory. If the *goal* of our ethical views is to

help us solve practical problems, we can easily see why “virtuous” ones will do this better. They are easier to understand and apply, and provide a better basis for social co-ordination. So far as we do seek theoretical virtue, the best explanation of why this makes sense comes from the pragmatist view that ethics is invented to solve practical problems.

Thus my argument in this paper is parallel to that given in the first. Many philosophers think that the view that ethics is invented to solve our problems makes a nonsense of the forms of moral inquiry we engage in. If we are not seeking purely moral truths, investigation could not bring about moral improvement, but would be “mere change”. But the very reverse is true. Moral inquiry, as it is actually conducted, would be pointless if it sought purely normative truths. Rather, we can explain why it is worth engaging in investigation, why the kinds of theories that we look for in inquiring are a genuine improvement on those that they replace, by appealing to a pragmatic theory. Neither practical norms nor moral inquiry can be grounded in realism and rationalism - they make better sense when we understand moral and practical thought as a process of inventing norms to solve practical problems.

My strategy of inversion is again on display in my third and fourth papers, which I conceive as a pair. One of the most distressing charges levelled at those - like me - who deny mind-independent truth and objectivity to ethics is that our position is somehow *immoral*. To show a proper appreciation and respect for moral normativity, then objectivity and truth are, as Dworkin claimed, something “you’d better believe” in. (Dworkin 1996) The assumption is that if cognitive knowledge, mind-independence, objectivity and all the rest are denied, then ethics must be somehow second-rate, lacking in authority. According to Parfit, denying these realist doctrines amounts to an affirmation of moral nihilism, the claim that nothing matters. (Parfit 2011) And that *would* be a terrible thing to think.

A traditional response has attempted to insulate anti-realism from such moralistic critique. Such metaethical claims, it is asserted, make no difference to the world of first-order, normative ethics, and as such cannot be assessed on moral grounds. But I deny this. Our metaethical theories about what counts as correctness and improvement in ethics affect our standards of moral inquiry - they determine the grounds on which we will be prepared to reject and revised the moral views that we started with. Since moral inquiry is an ongoing project, this means that whatever we believe in metaethics will indeed shape our normative ethical views.

But insofar as metaethical theories have these practical implications, they cannot, I argue, be established on purely theoretical grounds. We cannot be forced to accept any account of moral correctness simply by appeal to metaphysical or linguistic descriptions. Rather, it is a substantively normative question what account we should adopt of the standard of correctness in ethics. So the problem with the kind of attack launched by Dworkin and Parfit is not that moralistic critique of metaethical theories is *inapt*. The problems is that it is substantively *wrong*. Indeed, it has got things entirely the wrong way around. It is moral realism that is morally unacceptable, and pragmatic sentimentalism that can explain why morality matters.

I make the former case in “Immoral Realism.” It is precisely in accepting the inference from anti-realism to nihilism, made by Parfit and Dworkin - and at least implicitly, I claim, by other non-naturalists - that is morally objectionable.

As well as immediately world-directed moral judgements, such as the judgement that I ought to comfort my partner who is in distress, our moral viewpoints also contain higher-order, conditional commitments about the circumstances in which we ought to *change* our moral views. When we assess one another morally, we can criticise others for their higher-order judgements just as much as the

world-directed ones. As I argue, we should see metaethical claims simply as the most general and abstract of the higher-order conditions we place on our moral commitments.

Seen in this light, we can interpret non-naturalist realism in ethics as the claim that all our other moral judgements should only be accepted so long as they correspond to non-natural, purely normative, *sui generis* truthmakers. If there are no such truthmakers to be found - if naturalism is true - then we have no reason to retain any moral commitments. That's not to say that we ought to behave wickedly - it's simply that we should see morality as groundless, as lacking in authority.

But this is a terrible thing to say. There might be some general facts about the world that would constitute a reasonable basis for abandoning moral commitment - for example, if there were no external world or no other minds. But the fact that naturalism is true, that the world does not contain any *sui generis* normative facts or properties, would be a terrible reason to deny authority to morality. The truth or falsity of naturalism, the existence of such *recherché* entities as purely normative truths, would make no difference to the world we experience, or the history I share with my partner. It would be a betrayal to her and to others if I thought that my commitments would somehow be groundless in a natural world. Of course, realist philosophers *do* think that the world contains truthmakers for their moral claims. But it is the conditional normative claim that I object to, that moral commitments would not have authority if there were no non-natural truths.

No doubt most non-naturalists, if they discovered that naturalism were true and there were no *sui generis* moral facts, would not become moral nihilists, and would rather adopt a different metaethical theory. But they cannot, consistent with their own theory of normativity, explain *why* they ought to do this. And yet they need precisely such an explanation, for moral nihilism is morally unacceptable. The problem is not that, standing *outside* morality, the realist cannot explain why we should be

moral. No argument, I claim, can do that; nor should it aim to. The problem is that, standing *within* the moral perspective, the non-naturalist endorses the wrong conditions for abandoning morality.

Similar things could be said for other realist theories, although I do not go into detail in exploring these arguments in the paper. For example, Cornell realists like Boyd think that there are only moral truths if our moral languages refer to homeostatic property bundles. (Boyd 1988) But if the properties that we refer to in moral thought and deliberation do not exhibit this kind of homeostasis, then, according to this theory, there are no moral truths. Again, the fact that the world contained no homeostatic property bundles that serve as referents for moral talk and thought just would not be a good reason to abandon morality. Of course, as before, no doubt the Cornell Realist would not want to endorse nihilism in this scenario, and would instead prefer to adopt a different metaethical theory. And yet, given their theory of what moral correctness *is* - correspondence to homeostatic property bundles that are referents of current normative talk - they cannot explain *why* they ought to adopt a different view in this scenario. A good general test for any metaethical theory of moral correctness is whether it can provide a morally compelling account of what the world would have to be like for ethics to be baseless.

The final paper, "Sentimentalist Pragmatism Defended" turns to my own positive view about the standard of correctness in ethics. We should see ethics as marking out solutions to the collective practical problems which mutually sympathetic affective agents find in living together. Ethical claims are correct to the extent that they indeed provide solutions to these problems. Although I do not explore this issue in the paper, I think that such a view does answer the challenge just mentioned - it gives a compelling story of the conditions in which ethics would be entirely groundless. If there were no humans or other creatures that cared about things, or if humans failed to live in any kind of mutually engaged society, living in isolation or in utter indifference to one another, or if natural resources and human benevolence were so abundant that social life created no problems (as

imagined in the third section of Hume's second *Enquiry*), then I do indeed think that there would be no need for *ethical* norms. They would have no useful role to play. Furthermore, I think that this view gives the best possible *normative* answer to the question of why we should grant authority to ethics - because ethics helps us solve these problems that beset us. That is a far better response than simply pointing to the diktats of some mind-independent list of prescriptions, or urging the subordination of our human viewpoints to an "absolute perspective".

But my goal in this paper is not, primarily, to motivate or develop this view, but to respond to the charges of nihilism levelled by Dworkin, Parfit and others. After all, just because I think that there are moral objections to moral realism, it does not follow that anti-realism is not equally objectionable. And my view dispenses with so much that has been seen as key to grounding the normative authority of ethics. I see ethical claims as attempting to mark out solutions to problems, not corresponding with mind-independent, purely normative truths (although there are of course truths about what moral arrangements provide a solution to a given problem). Moral correctness is defined in terms of a particular set of problems. To us, these are felt as important and worthy of solution, but there is no guarantee that this would be seen as significant when viewed objectively "from nowhere". So my account denies that ethics has "perspectival" objectivity. And there might even be rational agents who cannot be persuaded to care about these problems, and hence cannot be motivated to obey the precepts of ethics. So ethics may lack rational sanction. Even if all humans can in principle be so persuaded, there is no guarantee that they will converge on one universal, timeless set of moral norms - our problems are highly contingent, and may admit of multiple equally good solutions. So there may be no timeless, universal set of moral norms to which we should grant authority.

I argue that none of these should be seen as necessary for us to acknowledge the authority of ethics. Morality could well have the sanction of an objective perspective, it might be compelling to all

rational beings, and moral inquiry may be fated to converge on a single point. But maybe not - and I am inclined to doubt any of these speculations. If I am right, then it is true that ethics is, in a sense, local - it is the kind of practical and intellectual activity that only creatures like us would see a point in. And it would mean that moral philosophy has less power than some might dream of attributing to it - there could be agents whom even the most sophisticated of ethical arguments would not persuade to care about morality. And moral philosophy may never find a unitary endpoint, continuing forever to invent new and diverse moral norms for the situations humans find themselves in. But we should not really care about these things - they in no way impugn, undermine, or denigrate morality. As I wrote at the start, I celebrate the mess, the human perspective, and the contingency of morality. They show why it is worth caring about.

The papers that follow do not take the form of a monograph; this is why I have not referred to them as chapters. I have been lucky to be offered the option of a “three-paper” dissertation (even if those three papers have become four) - lucky both because of the professional advantages of devoting my graduate career to writing journal-length papers, and because this form has offered the best vehicle for advancing this stage of my philosophical project. As such, each paper could be read alone, although I have tried to make the connections clear. And they could in principle be read in any order - they are presented in the order in which they were written, and, to an extent, in the order that seems to me to make most sense. And whilst they all share a common viewpoint and set of goals, they do not aim to present a single, fully-fleshed out philosophical position. Rather, I see them as clearing the way for the formulation of such a position.

Bibliography

Blackburn, Simon (2001) *Ruling Passions* (Oxford, Clarendon Press)

Boyd, Richard (1988) “How to be a Moral Realist” in G. Sayre-McCord ed., *Essays on Moral Realism* (Ithaca NY: Cornell University Press)

Dworkin, Ronald (1996) “Objectivity and truth: You'd better believe it” in *Philosophy and Public Affairs* 25 (2):87–139 (1996)

Parfit, Derek (2011) *On What Matters* (Oxford: Oxford University Press)

Paper One

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."

Introduction

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.

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 -

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.”

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 instances 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

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.

“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 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.

⁴ Peirce (1868) makes a similar point.

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.

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*

⁵ See especially Grahek (2007).

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 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

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.

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 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

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.

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.”

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.

(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 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.

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

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.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 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 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.

¹² Finlay (2009) makes a similar point.

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.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

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

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 *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 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.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 criticise "irrational" agents, but

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)

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. Lily 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 Lily'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, section II)

James recognises 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 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,¹⁵ 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 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.

¹⁵ 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.

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 (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)

Paper Two

Nonnaturalist moral realism and the limits of rational reflection

Abstract

This essay develops the epistemic challenge to Nonnaturalist Moral Realism. While evolutionary considerations do not support the strongest claims made by “debunkers”, they do provide the basis for an inductive argument that our moral dispositions and starting beliefs are at best partially reliable. So we need some method for separating truth from falsity. Many nonnaturalists think rational reflection can play this role. But rational reflection cannot be expected to bring us to truth even from reasonably accurate starting points. Reflection selects views that are coherent and conflict-free, yet there is no reason to think the nonnatural moral truth must be like this. Inasmuch as we seek coherent, conflict-free ethical viewpoints, that suggests our goal is not nonnatural truth at all.

Introduction

Nonnaturalist Realists in Ethics famously face an epistemic challenge. This paper argues that they cannot escape it.

Nonnatural facts, properties or truths are supposedly non-causal. This seems to rule out the possibility that some putative quasi-perceptual moral insight or “rational intuition” might reveal Nonnatural truths to us. Indeed, causal pressures which undeniably *have* had an influence in shaping our moral views and dispositions - those of biological and cultural evolution - seem to have nothing to do with Nonnatural Moral Truth at all. Against these worries, Nonnaturalists argue that evolutionary forces are not the *only* influence shaping our moral views. Some of our moral judgements - those offered by philosophers - are also the products of rational reflection and scrutiny, and this is a reason to trust them.

One response to this doubles down on the “debunking” power of evolutionary considerations. The rational methods Nonnaturalists appeal to involve reasoning from a starting set of judgements. We revise our opinions by using some beliefs to evaluate others, checking for consistency between individual judgements, and searching for greater coherence and systematicity among our belief-set

as a whole.¹⁶ Such procedures clearly cannot bring us to the truth when applied to a starting set of beliefs that is mostly false. The debunkers argue that evolutionary considerations show that our moral beliefs are, indeed, probably mostly false. So there is no material for rational reflection to work with.

Many philosophers recently have argued that debunking arguments cannot establish such a strong conclusion. I explain that this is correct - the debunking argument only succeeds if we either set the bar for epistemic justification *in general* so high that we must accept scepticism across the board, or accept a discredited account of scientific explanation. Still, this does not end the debate. As I show, the same evolutionary and etiological considerations can be marshalled to create a new, more modest argument, in the form of a pessimistic meta-induction. Our starting points might not be entirely or mostly false, but we must still conclude that they are at best a mix of truth and error.

This, I argue, is all we need to sustain the epistemic objection to Nonnaturalism, because rational reflection shouldn't be expected to guide us to truth in ethics, *even from somewhat correct starting points*. That's because there is no particular relation in which the ethical truths, as the Nonnaturalist portrays them, must stand to one another. There's no reason to expect that all correct ethical views will be assessed as "good" from the perspective of other correct ethical views. Although beliefs that entail logical contradictions cannot both be true, it's possible to believe almost any two ethical principles without inferring a logical contradiction. The most common type of incoherence between ethical views is *practical conflict*. But the truth might contain conflict. Conflicts may detract from the theoretical virtues of a theory - their simplicity, elegance, systematicity and so on. But there is no reason to think that the truth, as construed by the Nonnaturalist, will be simple, elegant and systematic. Some ethicists think that the moral truth *must* have these features; but there is no non-

¹⁶ Sometimes "Reflective Equilibrium" designates all such methods; other times, "Reflective Equilibrium" denotes a more specific process of working back and forth between particular and general judgements to bring them into line. Thus I use "rational reflection" as a catch-all term.

question-begging argument for why these assumptions should be more reliable than any of the other ethical judgements ethicists might reject or revise on the strength of them. Thus Nonnaturalists should assume that rationally-formed theories are no better than those derived from instinct or tradition. This is not quite scepticism, but it offers little comfort, considering how often we have judged the deliverances of instinct and tradition to be false.

But it's hard to deny that revising our ethical views to make them more coherent and conflict-free has nevertheless made them *better*. I suggest that this sense of betterness has nothing to do with *truth*, realistically construed. If we are moved to continue revising our theories to make them better in this sense, it shows that truth, Realistically construed, may not be our goal after all. Rather, as expressivists, pragmatists and constructivists suppose, our goal is non-alethic: "better" ethical theories are those which are useful, agreeable, or rationally acceptable.

One note going forward. Evolutionary arguments primarily target Nonnaturalist Realism; but there is debate as to whether they extend to Naturalist Realism (Barkhausen 2016, Street 2006) or Quasi-Realism (Street 2011, Blackburn ms). To assess whether my arguments generalise goes beyond the scope of this essay. So I offer a challenge, rather than an objection. Some Quasi-Realist and Naturalist theories resemble Nonnaturalism quite closely, such as Toppinen's quasi-realism (Toppinen, forthcoming) or Cornell Realism. These theories must answer two questions. Given the history of moral inquiry, why should they not accept metainductive pessimism about our moral starting points? And why think the "moral truth" must be coherent and conflict-free? Without answers to these questions, the suggestion remains that moral theorising does not aim at truth in *any* sense.

1. What do evolutionary arguments really show?

1.1 Reasoning from Error

Street argues, on evolutionary grounds, that our “basic evaluative tendencies” are probably wildly off-track with respect to the Nonnatural Moral Truth. But many philosophers (e.g. Parfit 2011, Scanlon 2014) respond that the countervailing influence of rational reflection on the formation of our considered moral views gives us a reason to trust them. The influence of reason in the genealogy of morality is a vindictory counterbalance to the influence of evolution.

Street denies that rational reflection has this power:

The objection gains its plausibility by suggesting that rational reflection provides some means of standing apart from our evaluative judgements, sorting through them, and gradually separating out the true ones from the false as if with the aid of some uncontaminated tool. [...] If the fund of evaluative judgements with which human reflection began was thoroughly contaminated with illegitimate influence...then the tools of rational reflection were equally contaminated. [...] Reflection of this kind isn't going to get one any closer to evaluative truth, any more than sorting through contaminated materials with contaminated tools is going to get one closer to purity. (Street 2006 p124)

This might seem odd - how could *rationality* be a “contaminated tool”? In fact, Street isn't trying to motivate evolutionary scepticism about rationality *in general*. It's just that rational reflection uses some of our judgements as a standpoint to critique others, and so cannot take us to the truth if not given true starting points as inputs. If our starting points are “likely to be false,” then rational reflection is no more than:

...a process of assessing evaluative judgements that are mostly off the mark in terms of others that are mostly off the mark. (p124)

Street's point is that rational reflection cannot turn muck into gold. Reflective Equilibrium and other methods of rational reflection cannot lead us to truth from mostly false views. Her argument thus hinges on her having *already* established that most of our views are likely to be false.

There are different ways to read Street's argument that our moral beliefs are “mostly off the mark.” But it is best interpreted as formally distinct from familiar, general sceptical arguments. The goal of Street's paper is to motivate scepticism *specifically* about moral truth, realistically construed, in such a

way that standard responses to (for example) external-world scepticism cannot simply be adopted in response. Many responses to scepticism argue that we can refute the sceptic if we are permitted to view our starting beliefs as at least defeasibly justified. “Debunking” arguments, of which Street’s is one, aim to show that *even if* we defeasibly assume our current beliefs as a starting point, we can generate an argument from claims we currently believe to undermine our views in the target area. After all, Street’s argument *relies* on the assumption that our beliefs about the genealogy of morality are broadly correct.

As such, Street’s opponent cannot be required to show how we would have had true moral beliefs in any conceivable world - such a story is impossible even with regard to everyday perceptual beliefs (or our beliefs about the genealogy of morality). The interest is to show that we would not have reliable moral beliefs *even if the world is largely as we suppose*. But this must extend to our suppositions about the *moral* world - just as we can only counter external-world scepticism, and explain how humans got to have accurate perceptual beliefs, by (defeasibly) assuming many of our current perceptual beliefs about the external world, so presumably Street’s opponent can (defeasibly) take her current moral beliefs as a defeasible starting assumption.

1.2 The improbable and the inexplicable

Enoch (2010) interprets Street as arguing that the supposition that we have attained the moral truth entails something “unbelievable”, and is hence itself unbelievable. As Street claims, if the genealogy of our moral beliefs nowhere makes reference to their *truth* as an explanatory supposition (which it couldn’t, since moral truth, according to Nonnaturalists, is causally inefficacious) then it would be an “unexplained coincidence” that the genealogy had nevertheless led us to the truth. But, as White (2010) points out, coincidences happen, and we are entitled to believe that they have happened when we have evidence that their results are instantiated. And we do seem to have evidence that we have come to the moral truth - whatever evidence we have for our moral beliefs themselves is, *ipso*

facto, evidence for the occurrence of a massive coincidence. Of course, this evidence is defeasible - *if we could not explain* how we came to the moral truth, then we would do better to adopt a sceptical attitude. So the real question is whether an explanation is possible.

Some philosophers have supposed that a result's being a massive coincidence implies that it is inexplicable. Coincidences are antecedently improbable, and these philosophers hold that explanation of P requires us to show why P was probable. For example, it seems antecedently improbable, given the starting physical conditions of the Universe, that conscious life would have emerged. Thus Nagel (2012) posits teleological laws to explain how consciousness arose. But almost all philosophers of science agree that the explanation of P *doesn't* have to show how P was antecedently probable. Explanation and prediction are asymmetrical. Unprotected sex with an HIV positive partner doesn't make infection probable (transmission rates are low) but it does *explain* infection. (Jeffrey 1969; Salmon 1971)

If explaining how we got true moral beliefs doesn't require showing how it was antecedently probable, the task is easier. As argued above, anti-sceptics don't need to provide an explanation for how we would have got to the truth in *any conceivable world*. We explain the emergence of reliable visual faculties by appeal to evolutionary advantage. But that only works in a world like ours, where visible properties of the environment have some relevance to survival (if we were brains in vats the explanation wouldn't work). Such an explanation for the reliability of our visual faculties, based on the assumption that ours is not a vat-world, is rightly taken to vindicate our beliefs about the external world.

It's true that no corresponding story can be given for the emergence of a moral-perceptual faculty specifically tailored for detecting moral facts - moral truths *as such* have no relevance to survival. Even if such a faculty were possible (which is dubious, given the causal inefficacy of the Nonnatural)

it is hard to explain why evolution would have endowed us with it. But nevertheless, as Joyce (2006) has argued at length, we can give an adaptive evolutionary explanation for the emergence, in our primate ancestors, of the psychological dispositions, including altruism and mutualism, that undergird the formation of pro-social societal norms, and the tendency to respond to violations of such norms with *moralising* reactions.

This is *not* like a perceptual mechanism for tracking the moral truth. According to Nonnaturalists, moral facts are not identical with any natural facts, hence facts about violations of pro-social norms are not *themselves* moral facts. The moral facts at best *supervene* on these facts. But that is enough to offer an explanation for our having at least some accurate moral beliefs. If the *moral* world is anything like what we think it is, there is a relationship between pro-sociality, altruism and mutualism, and the realm of the moral considered as such. Not every violation of altruism need be morally bad, nor need every moral fact supervene on facts about pro-sociality, altruism or mutualism. But if we explain the reliability of our perceptual systems by assuming that the world is roughly like what our perceptual faculties have revealed to us, then we can, likewise, explain the at-least-partial reliability of our moralising reactions by assuming that we live a world in which there is at least a fairly robust supervenience relationship between the pro-social and the moral. A world in which there is *no* connection between the pro-social and the moral is as distant from us as a world of vats, wires, brains and simulation software. In *our* world, evolution *itself* explains why at least some of our moral beliefs are probably true.

1.3 From Debunking to Pessimistic Meta-induction

Even if the debunking argument does not succeed in motivating thoroughgoing moral scepticism, the response above warrants only a limited optimism. It's explicable how the genealogy of morality could have endowed us with *some* correct evaluative dispositions. But we cannot suppose biological and cultural evolution gave us *uniformly* on-track dispositions, or even ones that are anywhere near as

reliable as our perceptual belief-formation mechanisms. That's not just because the moral truth did not casually regulate the progress of evolution. It's simply because we would expect natural selection to have also favoured certain dispositions that we take to be immoral. A predilection for global justice would have been disadvantageous for our ancestors; tendencies towards despicable nepotism would be evolutionarily advantageous. And, indeed, our ancestors don't seem to have had any commitment to global justice, and many of them accepted nepotism as legitimate. These examples are not exceptional. Many things that we take to be morally wrong would have been evolutionarily advantageous to believe, and indeed our ancestors had many wrong beliefs. Unless we can point to some countervailing force that would have brought us towards truth, we should, by induction, assume that we are in much the same position.

Of course, there must be some explanation of why our basic moralising dispositions are now being pressed into the service of securing global justice and other such goals. There are two immediate explanations. First, evolutionary explanations need not assume that all features of an organism are directly selected-for; selection happens at the level of suites of adaptations which come and go together, but not all of which are directly advantageous, and hence selected-for. The warmth of the polar bear's coat was selected for, but perhaps not its weight. The dispositions that brought us to value global justice or self-sacrifice for non-conspecifics may simply have come along with the useful adaptations that allowed us to live in mutualistic, norm-driven communities.

Secondly, we're not limited to *biological* evolution in explaining the emergence of our moral dispositions. Cultural evolution has also played a role. Perhaps it is this that pressed our basic pro-social dispositions into service in the promotion of goals like global justice. Indeed, cultural evolution also selects cultural adaptations, such as norms and belief-sets, at the level of suites - aspects of a culture's morality may simply have come along packaged together with things that were

directly selected for. It is not hard to think of examples among the strange practices and accretions that attend to many traditional moralities.

But while these two points help us explain the emergence of the moral beliefs we do have, they don't give us any reason to think them *true* - there's no basis for attributing reliability to these mechanisms. Indeed, there is good reason to think that cultural evolution often leads to false moral beliefs. As Barkhausen (2016) has argued, cultural evolution selects for moral principles and practices that are mutually advantageous between parties, making its outputs wildly contingent. Depending on the parties and circumstances, extremely different moral principles can be favoured. Since *extreme relativism* is something the nonnaturalist realist presumably denies, she must conclude that some of these principles will be off-track. Any non-relativist looking at the array of moral views actually in currency in different societies will conclude that cultural evolution sometimes leads people astray. Without further argument, we have no reason to think, no explanation of how it could have transpired, that our society was a special exception. Likewise, the manner in which adaptations are packaged in suites is totally contingent. Selection will sometimes favour non-advantageous adaptations that are pro-moral, but sometimes - perhaps often - not.

So we should assume that our starting points are probably riddled with error and our dispositions to make moral judgements are only partially reliable. Evolution might explain our having some true beliefs and reliable dispositions, but it cannot explain - in fact seems to rule out - our having uniformly or largely correct ones. We should expect to start with a pretty mixed bag. As such, whenever we take the influence of rational reflection to begin - be it now or in the distant past - we need to ask whether it could bring us (or have brought us) to truth from a starting point containing many falsehoods, since that is what biological and cultural evolution probably handed to us.

The argument just given is not the same as Street's debunking argument - it does not aim to undermine *all* our beliefs in a target domain. But it shares the following with debunking arguments: it begins by defeasibly assuming the correctness of our current beliefs. From that initial assumption, it argues that, in fact, it is wildly unlikely that *all* are true, although some may be. In this respect, it resembles the pessimistic meta-induction familiar from philosophy of science. In its classic form (Laudan 1981), the pessimistic meta-induction argues that the success of our theories cannot be a reliable indication of their truth, since successful theories have been found to be false in the past. That argument assumes the accuracy of previous assessments of falsehood, just as I do. However, while Laudan's claim that there have been many successful theories that are not even approximately true is controversial, it is surely clear that many of the moral beliefs that arose from cultural and biological evolution are outright false. The hypothesis that false moral beliefs often confer selective advantage seems very well-supported.

1.4 Intuition, instinct and plausibility

This is why our moral instincts cannot be trusted in the same way as our mathematical instincts. As Clarke-Doane (2012) points out, debunkers in Ethics make a mistake in assuming that the truth of Mathematical claims enters into the explanation of our having them. Mathematical truths aren't causally efficacious. Clarke-Doane argues that it would have been advantageous to believe that one lion and one lion number Two, even if, *per impossibile*, that weren't the case. But *in a world like ours*, we can see how beneficial it would have been to be disposed to at least some of the mathematical beliefs that are actually true. And it's very hard to think of *false* mathematical beliefs which it would have been beneficial, in this world, to think true. So, *given the starting assumption that the numerical facts supervene on the physical facts in largely the way we suppose*, we wouldn't expect any significant number of off-track mathematical dispositions to have been adaptive. It's not that evolution implanted the full extent of modern mathematical knowledge in our heads: rather, inasmuch as we do have some instinctual tendencies to find certain mathematical claims "intuitively" correct, there's no reason to suppose

that these will be systematically misleading. By contrast, as we have seen, in ethics we *should* expect to have some false beliefs and unreliable dispositions *even given the starting assumption that the moral facts supervene on the physical facts in largely the way we suppose*.

So our starting set of beliefs, prior to the application of rational reflection, probably contains many falsehoods. However, proponents of reflective equilibrium often claim that the process should be applied not to *all* our beliefs, but only to those antecedently selected as intuitively “plausible,” which count as our “considered judgements”. Scanlon claims that the

...force of the fact that we have arrived at certain judgments in reflective equilibrium depends on the substantive merits of the judgments we make along the way, in beginning with certain considered judgments and in modifying these judgments and others as we progress... (Scanlon 2014 p82)

But what are “considered judgements”? Scanlon continues:

One thing one needs to ask, in deciding whether something that seems true should be treated as a considered judgment, is whether it has any implausible implications or presuppositions. (p84)

Given what’s been said so far, should we expect intuitively “plausible” beliefs to be more likely to be correct than any others?

No. The arguments which show that our starting points are probably significantly erroneous are equally arguments that our intuitive assessments of plausibility are probably unreliable. The emergence of a set of reliable dispositions or a special faculty for intuitively detecting moral error seems highly implausible, given what we know about the biological and cultural genealogy of ethics. Indeed, such a capacity would be *counter*-adaptive, if, as argued, it is often adaptive to have false moral beliefs. A reliable capacity to intuitively find true beliefs more plausible than false ones would undermine the usefulness of false beliefs. So beliefs which survive direct assessments of intuitive plausibility are no more likely to be true than any others.

This should not be surprising. The judgements that a Christian fundamentalist or a medieval samurai find most “intuitively plausible” are no more likely to be correct than any of their other beliefs. Indeed, given the greater relative importance that their outlooks place upon doctrinal observance and honour (respectively), compared to altruism and equality, I would expect the judgements they take to be most plausible to be *less* likely correct than their average beliefs. After all, for a fundamentalist, the goodness of altruism is subordinate to the revealed will of God - if the Text proscribes altruism it must be rejected. For a samurai, altruism may similarly wait upon honour. If a fundamentalist or a samurai were only to reason from their most “intuitively plausible” moral beliefs, they might exclude the correct altruistic judgements that they share with us.

This contrasts with mathematics. Individual mathematical beliefs can come from poor-quality testimony, guesswork or faulty memory - of course beliefs from such sources are unreliable. But if the mathematical supervenes on the natural in the way we assume, we can explain the emergence of reliable, specifically mathematical dispositions. So our antecedent selection of “plausible” starting points in mathematics *should* be granted weight. Those that withstand direct scrutiny are more likely to be true than average, since these are presumably the products of our reliable mathematical dispositions, rather than memory, hearsay and guesswork.

If all this is correct, much rests on the competence of rational reflection to sort truth from falsehood in ethics - far more than in other domains. In our beliefs about the external natural world, we have the advantage of causal-perceptual inputs. We cannot have these in ethics. And in mathematics we can explain why our antecedent assessments of plausibility might be reliable. We cannot similarly explain why intuitive assessments of plausibility in ethics would be reliable. And in neither our beliefs about the external world nor about mathematics should we expect the influence of selective evolutionary pressures to be frequently falsehood-conducive. In ethics, we know that evolution has

frequently selected for the unethical. So in ethics there is surely significant work to be done in sorting truth from falsehood.

2. Why Rational Reflection cannot sort truth from falsity

2.1 Reflection as internal assessment

Can we expect rational reflection to weed out the false from the true? I claim there is no reason to think so, if it consists in no more than the familiar methods of reasoning from a starting set of judgements - using some beliefs to assess others, checking our belief-sets as a whole for consistency, or seeking theoretical virtues like coherence and systematicity.

Start with the simplest forms of rational reflection. Street gives two distinct characterisations of reflective equilibrium. One version consists in “assessing evaluative judgements ... in terms of others” (Street 2006 p124). We can see how this would operate. We can use an ethical principle or judgement as an “evaluative perspective”, a lens through which to view the world. When we do so, we determine whether the things we so view are good, not in the “all things considered” sense, but simply good *in terms of the value* through which we are viewing them. We can turn this gaze inward and see whether, from the perspective of one value we are looking through, another value we accept looks good. If it does not, we could revise or reject the “bad” value.

The problem is that there’s no reason to assume that all true value judgements will look good from the perspective of all other true value judgements. Seeming good from a true evaluative perspective obviously isn’t a criterion of truth in regular factual contexts - all sorts of regrettable things are true out in the world. But even within the evaluative sphere, we can see instances of values we accept as true, but which look unpleasant when viewed from other values. It’s true that liberty is important, but insistence on liberty looks unsavoury when viewed from a perspective that judges in terms of equality. It’s true that partiality towards our spouses and children is good, but this seems regrettable

when viewed from the more universalistic perspective of justice. Of course, all true principles will seem good from the “all things considered” perspective of someone who already knows the entire moral truth; but they needn’t seem good viewed from any more limited perspective.

2.2 Reflection as consistency

Street also describes Reflective Equilibrium as a process whereby we “test our evaluative judgements only by testing their consistency with our other evaluative judgements” (Street 2006 p124). If testing for consistency just means eliminating logical contradiction, then this seems obviously truth-tracking: after all, *pace* dialetheists, contradictions cannot be true. However, assessing value judgements for logical contradiction is not straightforward. Does the view that “we have a reason to maximise the good” *contradict* the view that “we have a reason not to lie”? Not obviously - any two reasons can co-exist. We cannot immediately infer the non-existence of one reason from the existence of another.

Normally, we take belief in one fact to contradict belief in another fact if we can infer a contradictory proposition from the two. We might think that the two principles just stated *do* entail contradictory propositions, and hence cannot both be true. For example, someone who only believed the former principle would be a consequentialist, and so might infer, in a given situation “I have a reason to lie, and no reason not to,” which of course contradicts the latter, deontic principle that “I have a reason not to lie.”

The problem is that practical judgements are inferentially non-monotonic. If I *only* believe the consequentialist principle, then I will make inferences that logically contradict the deontic principle. But if I accept both principles (in other words, I believe in consequentialism with “side constraints”), the inference is not the self-contradictory “I have a reason to lie, and I have no reason not to lie, and I have a reason not to lie”, but “I have a reason to lie, and a reason not to.” That principle A and

principle B would yield logically contradictory claims if I held each individually does not imply that the principles are logically contradictory; given the non-monotonicity of practical inference, I can believe both without inferring any contradiction. Thus, there is no *a priori* reason why I should reject one on the strength of the other, if my goal were to acquire the truth.

(Briefly: the logic of this is that one cannot infer $\partial[A \ \& \ \sim A]$ from $[\partial A \ \& \ \partial \sim A]$ (where ∂ ="ought" or, "there is a reason to"). Holding the latter claim does not violate the principle that "ought implies can" since, although I cannot perform $[A \ \& \ \sim A]$, I can perform each of A and $\sim A$.) (Williams 1973 Ch11; 1981 Ch5)

There's a significant question whether the concept of *obligation* in itself rules out this kind of situation - that there's always *one* thing that you *ought* to do (although of course many philosophers argue for the existence of "tragic choices"). But even if we think that outright *obligation* is always univocal, almost everyone accepts the possibility of *gradable* prescriptive claims, such as those concerning prima facie obligations, pro tanto rightness, or reasons for and against. Even if we think that I can never have an obligation to P and not-P, there doesn't seem to be anything conceptually impossible about a situation in which I have reasons to P and not to P, or where it is *pro tanto* right to P and not to P, and so on. When it comes to graded prescriptive and evaluative concepts, what might have looked like logical contradictions are simply *conflicts*.

This is not to say that we *never* have logically contradictory ethical beliefs. Sometimes the correct interpretation of a moral principle is that it involves a *negative* existential statement about what reasons¹⁷ exist, such that it directly contradicts another positive claim about what reasons exist. For example, some people who believe that "lying is wrong" don't just mean that "there are always reasons not to lie", but that "there are never any reasons whatsoever to lie." If many of our

¹⁷ Or *prima facie* obligations, or whatever.

principles were of this strident form, then there will be contradictions, and thus truth-directed reasons to revise our views: it just cannot be true that we *never* have any reason whatsoever to lie, and that we *always* have at least some reason to maximise the good. These things will generate strict contradictions in practice.

It's worth noting how strong such claims are. Even someone who thought it was always actually wrong to lie needn't accept that nothing ever counts in favour of lying. Most people who accept a prohibition on lying concede that there are sometimes considerations in favour of lying in certain situations - they simply insist that these are *outweighed* by the wrongness of lying. The starting beliefs we've been endowed with by biological and cultural evolution seem to be mostly about what is a reason for what, rather than strident negative claims about what reasons never exist. So it strikes me that simply purging our belief system of logical contradictions won't get us very far away from our starting points - which is worrying, if our starting points probably contain significant error.

It's true that even once we have deleted the contradictions, we may still have a viewpoint that is *conflict-ridden*. We will often believe that we have reasons - maybe powerful reasons - that pull in opposite directions. But surely no-one has any basis to think the Moral Truth cannot contain conflicting reasons. Although conflict-ridden moral outlooks are unpleasant to live with and hard to use in practical deliberation, there is nothing in the Realistic notion of mind-independent moral truths, of objective Reasons, that rules it out. Any two reasons can consistently co-exist. The Moral Truth may be ridden with conflicts between gradable prescriptions.

2.3 Reflection and theoretical virtue

Of course, many philosophers *have* sought to systematise their ethical viewpoints to avoid excessive conflict. For them, the presence of too many conflicts in a theory is a reason to abandon or revise that theory. This is not because conflicts are literally contradictory, and so not *possibly* true, but

rather, as Kagan (1989) explains, they detract from the *coherence* of the theory in a broader sense. Portraying rational reflection as a search for coherence in some sense that *goes beyond* mere logical consistency seems true to the classic description of Reflective Equilibrium as a process of “working back and forth” between our particular and general judgements in order to bring them into “equilibrium.” Indeed, the search for coherence in our ethical theories might be seen as of a piece with the common preference that researchers in many domains have for theories that exhibit “theoretical virtues” - theories that are simple, systematic, explanatory, and so on. Perhaps the broader notion of coherence sought in Reflective Equilibrium just *is* a matter of simplicity, systematicity and explanatoriness; or perhaps it stands alongside these as a virtue of ethical theories. Either way, we can understand rational reflection as the search for coherence and other theoretical virtues in our ethical outlooks.

But why think that the virtues - coherence and whatever else - are indicators of ethical truth? Ethicists frequently appeal to coherence and other virtues in theory selection, but they are less careful to explain why they assume that the ethical truth will exhibit the virtues, if indeed they do. There certainly doesn't seem to be any *a priori* reason to think that the moral truth, as portrayed by the Nonnaturalist, needs to be coherent or systematic. It seems entirely reasonable to me to imagine that the truth will be highly complex rather than simple (Griffin 2015) with a profusion of independent goods, requirements and virtues, and that it will be full of conflict, (Williams 1973 Ch11) with areas of moral indeterminacy (Scanlon 2014) and vagueness rather than prescriptions for every situation, and that individual and piecemeal judgments may not always be explained by deeper or more general principles. In other words, whatever coherence is, *over and above* non-contradiction, there is no reason to assume in advance that the truth will be coherent (or, alternatively, able to be represented by a coherent theory). And likewise for any other theoretical virtues. Perhaps this is just how things are.

I have not argued that moral truth *does* contain lots of conflicts, or that it falls short of theoretical virtue in any other way. My point is simply that there is no *a priori* reason to rule this out, and hence, no reason to expect any methodology of rational reflection that goes further than simply avoiding logical contradiction to bring us closer to the truth. All it takes for an ethical proposition to be true on the realist view is for it to correspond to a moral fact, and there is no limit on what moral facts can exist. Parfit and Scanlon and others might hope that the truth is not such as to generate conflicts. But it is hard to see what evidence they could have for this view. The views handed down to us by biological and cultural evolution manifestly *do* contain conflicts, and there is no *a priori* reason why these must indicate falsity.

If a philosopher insists that the moral truth *must* be coherent and conflict-free, she needs to offer some reason to think that *this* judgement is itself any more likely to be reliable than the array of other moral judgements she will reject on the strength of it. If our first-order starting points, the ground-level moral beliefs to which we apply reflective reasoning, are likely to be partly true and partly false, surely we must assume the same of our meta-moral beliefs about the structure of the moral truth. If we have no reason to think that our dispositions to make moral judgements and to find certain claims intuitively plausible would be highly reliable, then I think we have no reason to think that our dispositions to make judgements about the structure of the moral truth would be highly reliable. If the Nonnaturalist's epistemic challenge is to explain why rationally-formed beliefs are likely to be true, it strikes me as simply begging the question if her answer assumes the truth of a contentious and unobvious belief about the structure of the moral truth.

Indeed, if our starting point includes *both* a messy assortment of conflicting, unsystematised and poorly coherent atomic moral judgements, *and* the belief that the moral truth must be conflict-free, systematic and coherent, then surely the simplest way to eradicate the contradiction is to abandon the latter belief on the strength of the former set, rather than embarking on the huge task of

revising the former set to make it systematic and conflict-free. So, if anything, the most basic form of rational reflection - contradiction-eradication - undercuts the assumptions needed to legitimate the more revisionary forms that aim for conflict-eradication and systematisation. In this spirit, Griffin inveighs against the distortion caused by a “Newtonising” obsession with simplicity and systematicity (Griffin 2015). In a similar vein, Wood (2016) argues that many philosophers have been “ravished by the formal beauty” of “very elegant abstract formal theories” to end up with “shallow” views that are “revolting and inhuman”.

Philosophers appealing to theoretical virtue in theory-selection often support their methodology by pointing out that scientists do the same thing. It’s widely accepted that, due to the underdetermination of theory by evidence, appeals to theoretical virtue are indispensable in science if scientists are to be able to select unique hypotheses for acceptance. But it’s a matter of debate in philosophy of science whether the virtues play this role because they indicate truth or for some other reason. Philosophers since Bacon¹⁸ and Hume¹⁹ have reminded us that an excessive love of systematicity, simplicity and elegance can lead to false theorising. Certain kinds of theories appeal to us, but this is no reason to think the *world* must be like that. Many philosophers of science have denied that the virtues are well-correlated with truth at all - Levi (1997) argues that the virtues are actually negatively correlated with truth, but that virtues affect the balance of options in cognitive decision theory because virtue adds to the “epistemic utility” of beliefs. And even those philosophers who argue that considerations like simplicity are at least defeasibly truth-indicating are careful to point out that this is a contingent, contextual fact, as in Sober’s extensive examination of the virtue of simplicity (Sober 2015). As Sober points out, we should only expect simplicity to guide us to truth given the presence of a variety of quite specific background assumptions, which don’t always obtain.

¹⁸ “The human understanding, from its peculiar nature, easily supposes a greater degree of order and equality in things than it really finds.” Bacon (1905) Book 1 § XLV

¹⁹ “...that love of *simplicity* which has been the source of much false reasoning in philosophy.” Hume (1975) Second Enquiry, Appendix II § 250

If seeking simplicity has often brought us true scientific theories, we can argue inductively that the virtues are good indicators of truth *in the sciences*. But there's absolutely no reason to infer that this induction carries over into ethics.

2.4 Coherence as a practical value

One final role for coherence and the other virtues remains, but it offers no succour for the Nonnaturalist. Parfit (2011) argued that there *has to be* a single true ethical principle that captures all of morality. He thought it would be “a tragedy” if there were no one rule. But as we have already seen, this argument cannot show us what is *true* - the truth can be uncomfortable and even unpleasant. As Blackburn (2011) says “outside the charmed walls of All Souls College, there actually are tragedies”. But here lies the irony. Realists, and especially nonnaturalists, have long argued that truth in ethics is independent of whatever anyone happens to think about it; to call something “true” is to do something *over and above* endorsing, recommending or approving of it. This is *why* they are forced to accept that an ethical claim's seeming like a tragedy from the perspective of other of our ethical beliefs cannot count as evidence against its truth - the mere fact that we disapprove of some ethical belief has no bearing on its truth or falsity. One major realist criticism hurled at expressivists like Blackburn is that they allegedly cannot distinguish between calling an ethical claim “true” and simply approving of it.

But perhaps the expressivist claim, that endorsing and judging true are closely connected, better represents the actual methodologies we use in ethical theorising. While it's dubious that the lack of a single overarching ethical principle would really constitute a tragedy, there are ways that the ethical world *could* (in the nonnaturalist sense) be, that would, I think, be tragic. If ethical conflict were powerful and pervasive, constantly placing incommensurable demands upon us, or if there were huge areas of life where the ethical facts offered no guidance at all, or if ethics required that *fiat*

iusticia in a way that made it certain that *pereat mundus*²⁰, then I think that *would* be a tragedy. Furthermore, *because* I find such situations tragic, I am prepared to reject any moral theory that takes such forms. I suspect that such considerations are frequently employed in ethical theorising - we seek ethical theories that are not tragic, unpleasant or deeply uncongenial. If it is, then expressivists, who think that calling an ethical theory “true” just *is* approving of it, can do a much better job of explaining the way we theorise than Nonnaturalist realists, who deny the connection.

Similarly, Scanlon argues that reflective equilibrium’s systematising aim of “finding general principles that account for one’s beliefs” has “important benefits” (Scanlon 2014, p84). Certainly, theoretically virtuous ethical outlooks have benefits. Simple, coherent, mutually supportive, systematic, conflict-free theories are practically useful. They are an excellent basis for deliberation, discussion and establishing social co-ordination. Inasmuch as our views are explanatory, we can explain ourselves to one another. Inasmuch as they are simple, we can swiftly work out what to do. Inasmuch as they are conflict-free, we can live without the exhaustion of guilt and moral anxiety. Inasmuch as they are systematic, we can unite our evaluations in disparate spheres of life. But of course, there is nothing in the Nonnaturalist notion of ethical truth that makes us think that the correct ethical theory must be practically useful in this manner. It is ethical *pragmatists*, who see ethics as a “social technology” (Kitcher 2011; 2012) - a practical tool to serve collective human goals, dependent on our interests rather than an ultimate authority at whose feet we must bow, regardless of the cost - who can explain why we should prefer useful theories. Again, I suspect that many ethical theorists *do* seek practically useful theories - if so, that suggests that Pragmatism, not Nonnaturalism, is the metaethical picture that makes sense of our investigatory conduct.

It should not surprise us that Reflective Equilibrium and other methods of rational reflection make better sense against the backdrop of Anti-Realist than Nonnaturalist metaethical pictures. Reflective

²⁰ “fiat iusticia, pereat mundus” - “Let justice be done, though the world should perish.”

Equilibrium was introduced into ethical theorising by Rawls, who was a Kantian *constructivist*, not a Realist. For Kantian constructivists, acceptability to rational reflection is itself the standard of correctness; there's no *further* question as to why coherence - or whatever rational reflection seeks - will also lead to truth. Acceptability to rational agents as such is all there is to moral truth. Nonnaturalists reject this by definition - the moral truth is independent of whatever we or anyone else thinks, or is disposed to think.

Conclusion

We can draw two conclusions from this.

Evolutionary and other genealogical considerations don't force the Nonnaturalist into extreme moral scepticism as Street supposes. But they do show that any moral viewpoint that relies only on instinct and tradition is overwhelmingly likely to contain numerous errors. Since there is no reason to think rational reflection will bring us closer to the truth, as Nonnaturalistically conceived, even rationally-formed moral outlooks are probably also riddled with error. Rational reflection is no more reliable than tradition or instinct. This forces Nonnaturalists into a pessimistic anti-Theory in Ethics: the ethical systems of philosophers are no more likely to be true than those of anyone else, and all are full of error.

But *if* we find the methods of rational reflection compelling - if we think that coherent virtuous theories are *better*; if we find the consideration that our views are the products of rational scrutiny *vindicating* - then we should abandon Nonnaturalism. We should prefer theories that appear morally congenial, useful and rational only if we believed some non-realist kind of view - expressivism, pragmatism or constructivism. If we are not pessimistic anti-Theorists, we should not be non-naturalist realists.

Bibliography:

Bacon, Francis (1905) *Novum Organum* (New York: P.F. Collier)

Barkhausen, Max (2016) "Reductionist Moral Realism and the Contingency of Moral Evolution," *Ethics* 126 (3):662-689

Blackburn, Simon (2011) review of "On What Matters" accessed at <http://www2.phil.cam.ac.uk/~swb24/reviews/Parfitfinal.htm>

Blackburn, Simon (ms) "Sharon Street on The Independent Normative Truth as such" accessed at <http://www2.phil.cam.ac.uk/~swb24/PAPERS/Meanstreet.htm>

Enoch, David (2010) "The epistemological challenge to metanormative realism" *Philosophical Studies* 148 (3):413 - 438 (2010)

Griffin, James (2015) *What Can Philosophy Contribute to Ethics?* (Oxford: Oxford University Press)

Hume, David (1975) L. A. Selby-Bigge and P. H. Nidditch (eds.), *Enquiries concerning Human Understanding and concerning the Principles of Morals* (Oxford: Oxford University Press)

Jeffrey, Richard (1969) "Statistical explanation vs. statistical inference" *Nicholas Rescher (ed.), Essays in Honor of Carl G. Hempel. pp. 104--113* (Reidel)

Joyce, Richard (2006) *The Evolution of Morality* (MIT Press: Cambridge MA)

Kagan, Shelley (1989) *The Limits of Morality* (Oxford: Oxford University Press)

Kitcher, Philip (2011) *The Ethical Project* (Oxford: Oxford University Press)

Kitcher, Philip (2012) *Preludes to Pragmatism* (Oxford: Oxford University Press)

Laudan, Larry (1981) "A Confutation of Convergent Realism", *Philosophy of Science, Vol. 48, No. 1, (Mar. 1981): 19-49*

Levi, Isaac (1997) *The Covenant of Reason* (Cambridge: Cambridge University Press)

Nagel, Thomas (2012) *Mind & Cosmos* (Oxford: Oxford University Press)

Parfit, Derek (2011) *On What Matters* (Oxford: Oxford University Press)

Salmon, Wesley (1971) *Statistical Explanation and Statistical Relevance*

Scanlon, T.M. (2014) *Being Realistic About Reasons* (Oxford: Oxford University Press)

Sober, Elliot (2015) *Ockham's Razors* (Cambridge: Cambridge University Press)

Street, Sharon (2006) "A Darwinian Dilemma for Realist Theories of Value" *Philosophical Studies* (2006) 127:109166

Street, Sharon (2011) "Mind-Independence Without the Mystery: Why Quasi-Realists Can't Have It Both Ways" in *Oxford Studies in Metaethics* 6:1-32

Toppinen, Teemu (forthcoming) "Non-Naturalism gone Quasi" *Oxford Studies in Metaethics*

White, Roger (2010) "You Just Believe that Because..." *Philosophical Perspectives* 24 (1):573-615 (2010)

Williams, Bernard (1973) *Problems of the Self* (Cambridge: Cambridge University Press)

Williams, Bernard (1981) *Moral Luck* (Cambridge: Cambridge University Press)

Wood, Allen (2016) "Kant Marx Fichte," interview in 3AM <http://www.3ammagazine.com/3am/kant-marx-fichte/>

Paper Three

Abstract

Non-naturalists 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.

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 "judgement-independent" fact. According to a prominent version of Realism, supported by Derek Parfit, T. M. Scanlon, Thomas Nagel, David Enoch and many others, it's not a *natural* fact either. When we make certain moral claims, we are stating beliefs that correspond with irreducible, *sui generis*, purely normative facts. These philosophers are "Non-naturalist" (or non-reductive) realists.

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. 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 once argued that anti-realists like

Blackburn held something *immoral* when they denied objective truth to ethical claims. (Parfit 2011; Dworkin 1996)

Williams 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? (Williams 2006) 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 are 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 moral realism. Realists make us *conditionalise* our world-directed moral commitments on the wrong things. Question 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.

Conditionalisation

We have all sorts of moral beliefs whose objects are located in the world around us. I may think that I ought to comfort my partner if she is in pain, or to keep the promise I made to my friend that I would carefully proof-read this paper before submitting it to a journal.

But to understand the totality of a person's moral outlook, it's not enough just to know what their directly world-directed moral attitudes and beliefs are. None of us hold every aspect of our moral outlook to be entirely unrevisable. There are circumstances in which we think we ought to change our views. As a result, we also have higher-order moral judgements about the conditions in which we ought to change our world-directed moral judgements or attitudes.

These form a major part of everyday morality. I have a special duty to my partner, to, for example, comfort her when she is 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 five years. Of course, unless I'm unreasonably jealous and suspicious, this thought won't enter into my everyday moral deliberations. That's why it's best to think of the conditionalisation as a higher-order belief or attitude, rather than as a proper constituent of my everyday, partner-directed beliefs and attitudes.

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 moral beliefs and attitudes regarding our partners, may accept the same duties of other-regard and care on a day-to-day basis. But if you conditionalise these, not just on fidelity, but also on your partner maintaining a certain 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 direct, 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 - cases where we should radically revise our moral views, or drop our commitment to morality altogether. If there is no external world, and all the other people I think I interact with are mere figments of my imagination, leaving me as the only conscious agent, then it seems right to have no world-directed moral beliefs at all, and concern myself only with narrow self-interest. Moral nihilism (if not prudential nihilism) seems plausibly appropriate in this situation. But there are very few other cases where we feel we ought to abandon morality altogether. In other cases we *must not* become nihilists. Some conditionalisations are wholly reprehensible.

What is Non-Naturalist Realism?

Why do I think that Non-naturalist realism forces us to conditionalise moral commitment on something objectionable? There are, of course, as many formulations of non-naturalist realism as there are non-naturalist realists, and it is not my intention to anatomise every extant view in the area. 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.

A) 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 in *some* sense to call certain ethical claims “true” - 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.

Dummett, who coined “anti-realism” as a philosophical term of art, characterised realism and anti-realism in the following terms:

Realism I characterise as the belief that statements of the disputed class possess an objective truth-value, independently of our means of knowing it: they are true or false in virtue of a reality existing independently of us. The anti-realist opposes to this the view that statements of the disputed class are to be understood only by reference to the sort of thing which we count as evidence for a statement of that class. (Dummett 1978 p.146)

On this view, the defining feature of realism is the claim that moral truths possess a truth-value independent of whatever we happen to think or feel. This seems entirely right. Nagel writes: “Normative realism is the view that propositions about what gives us reasons for action can be true or false independently of how things appear to us.” (Nagel 1973 p139). By contrast, for Dummett the defining feature of anti-realism is its understanding of correctness in terms of a procedural epistemic notion, portraying moral truth in terms of moral evidence.

But it is not clear that these are contraries, and philosophers since Dummett have attempted to bridge the divide. For example, Scanlon argues that we can preserve the idea that moral facts are “independent of us” without providing any metaphysically heavyweight account of a moral *reality*. All we need is a suitable epistemic story of how to identify correctness and incorrectness in the normative domain; that’s enough to vindicate realism so long as that epistemic story allows for “the possibility that some facts about the subject may outrun our ability to discover them.” (p71) In other words, Scanlon both asserts that moral truths are independent of us and of what we happen to think, and claims to characterise moral truth in terms of moral evidence.

It is unclear to me that Scanlon’s attempt to bridge Dummett’s divide, as presented in his (2014), succeeds. However, for present purposes I leave it as an open question whether such an account

might be developed. Given this, I think it best to focus solely on the issue of *independence*²¹ as the criterion of realism, and drop Dummett's definition of anti-realism. Realists don't just think that some moral claims are true - some moral claims would be true *regardless of what anyone happened to think*. Whether this claim must be given a metaphysical construal, or can indeed be accounted for in purely epistemic terms as Scanlon suggests, it certainly has epistemic implications. For it explains the possibility of a certain kind of error. Any theory that allows us to call some moral claims "true" will also, presumably, allow us to designate others "false", but the non-realist theories listed above struggle to account for the possibility of widespread error. However, if we hold that moral claims have their truth value entirely independently of the judgements of moral investigators, we can see how some people, even entire communities and cultures - can have got things wrong. Perhaps, once upon a time, the entire human race was greatly morally mistaken.

When people get things wrong, according to the non-naturalist, 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 "just" 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. Of course, by symmetry, this means that we, too, could in principle be wrong. In his characterisation of realism in general, Nagel claimed that "Realism makes scepticism intelligible." (Nagel 1986 p90) As I have portrayed the contours of the debate, that is exactly right as regards moral realism.

B) Non-naturalism

²¹ 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 "judgement-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 judgements, stances and responses of humans. For convenience, I will talk mainly of "independence" or "judgement-independence."

What makes judgement-independent facts *non-natural*? Some “naturalist” realists think that there are judgement-independent moral facts, because moral concepts refer to natural facts, which are judgement-independent. These are discovered through empirical investigations. Of course, no-one doubts that natural facts, discovered through empirical investigations, are highly relevant to moral judgements. But, object non-naturalists, couldn't some people have the *wrong* moral concepts? There has to be some deeper fact about *which* natural facts matter morally. And such deeper facts *couldn't* be discovered through empirical investigations, they say - these are *non-natural* facts.

What distinguishes these two kinds of facts? Again, the literature here is complex. But one issues that almost all non-naturalists agree on is that irreducible normative facts are *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 p493)

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 p85)

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. But divinities are paradigm cases of entities that are not natural. And this portrayal can be agnostic as to whether all natural facts are causal. Rather, the claim is more minimally, that, so long as normative facts are understood as non-natural, they are understood as non-causal.

A few non-naturalists have bridled against the assumption that irreducibly normative, non-natural facts and properties cannot be causes. In his review of Parfit, Larmore writes:

It is not at all obvious that reasons cannot be causes...not reasons psychologically conceived as belief–desire pairs (as by Donald Davidson), but normative reasons objectively conceived. (Larmore 2013)

So whilst it is true that most non-naturalists believe that non-natural normative facts are *not* causal, it takes more intellectual work to show that they *cannot* be understood as causal.

But, even absent this argument, it is not surprising that most non-naturalists think that non-natural normative facts are not causal. They think this because they insist that these facts, or at least a highly significant subclass of these facts, are not susceptible of empirical detection. Scanlon frequently contrasts “empirical” discoveries and judgements with normative (and mathematical) ones. Parfit writes:

The most fundamental normative facts are not...empirically discoverable facts. (Parfit 2011 Vol.II p307)

It is plausible to think that tractability to empirical detection is a sufficient condition for something to count as natural.²² If normative facts *did* enter into causal relations, we would need some explanation of why they are not open to empirical detection. Hence non-natural facts are most plausibly seen as non-causal.

This obviously raises epistemological questions, and much of the literature on this topic attempts to explain *how* normative facts are to be discovered if they are non-causal.²³ Non-naturalists insist that a plausible epistemology for non-causal subjects can be given, and that the case of morality is no more puzzling than knowledge of mathematical, logical or modal facts. But that is not my subject here.

²² Arguably not necessary, depending on how we understand the modality of the notion of “tractability”. There may be some objects in the universe that are moving away from *us* at such a great speed that we can never receive information about them. So we will never empirically detect them.

²³ 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 Ch32), Scanlon (2014 Ch4), Enoch (2011) and Clarke-Doane (2016).

If non-natural facts are non-causal, this means that they make no difference to the world of our experience. Every natural fact about my history with my partner, our interactions and our shared understandings would be just the same, whether or not I really did have obligations to her.

So, when non-naturalists make the Realist claim that people could be wrong in ethics, they don't just mean that they could be wrong about the natural facts. For them, not only does scepticism make sense in principle, but this kind of scepticism is *independent* from scepticism about the natural world. The facts about which we can thus be sceptical have no causal impact on us, our experiences, or anything else.

C) 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 judgements 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 p46)

In a similar vein, Matthew Bedke, claims:

True, substantive reason propositions require an ontology that the natural, or the natural cum phenomenal, does not afford. (Bedke 2012 p128)

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

On such [Naturalistic] views, there aren't really any normative reasons. (Parfit 2011 Vol.I p110)

elsewhere:

Naturalism and Non-cognitivism are both, I shall argue, close to Nihilism. Normativity is either an illusion, or involves irreducibly normative facts. (Parfit 2011 Vol.II p267)

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 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, judgement-independent truths that it counts as genuinely normative. If they think that, they agree with Parfit's claim that, if there are no truths of the relevant sort, moral talk and practice lacks genuine normative force. Nothing matters.

What is Non-naturalist Realism: Summary

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

A1) Some moral facts are independent of what anyone happens to think.

—>A2) Moral scepticism at least makes sense.

B1) Moral facts are causally inert.

—> B2) Moral facts make no difference to the experienced world.

C1) Only non-naturalist realism explains how things matter in the normative sense.

—> C2) If there are no facts of the sort defined in A&B, nothing matters in the normative sense.

Four 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. It's quite acceptable to adopt the conditional belief that, if we were to discover that there is no external world, or that it had none of the relevant objects of moral concern, then I would absolve myself of all moral requirements. 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. It scorns them if I am prepared to

abandon my sense that there are some things I may, and some things I must not, do to these people, just because I find out there is no God.²⁴

3) The Faithless conditional

One way of making vivid what's wrong with the two above conditionals is to return to our original example of partner obligation. If I conclude that nothing matters, then I conclude that I really don't have any moral reason to comfort my partner when she is in distress. If I conclude this just because I have become an atheist or seen the Mets lose, then I betray my relationship with her.

That's not to say that nothing should bring me to suspend or alter my commitments to my partner. As mentioned before, if she has been cheating on me, then a change appears in order. Perhaps if it turned out that the degree of my partiality to her was causing excessive suffering to others then I might reasonably come to rethink the extent of my partial obligations. In other words, conditionalising my commitments to my partner is not, in itself, wrong. The issue is the content of the conditionalisation. The questions of whether the Mets win, or whether God exists, are just are not the kinds of things my obligations to her should be conditionalised on.

4) Parfit's conditional

Derek Parfit said that, if Naturalism is true, then nothing matters. I think it 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 it simply follows from non-naturalist realism.

²⁴ It is quite possible to believe in a God without endorsing Ivan's conditional. William James was a theist. Yet he writes:

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 p198)

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." In that sense, James' ethics is fundamentally humanistic; it does not conditionalise moral commitment on anything that I object to.

As stated in C1, 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 A2, moral scepticism makes sense - the question of what moral facts there are is entirely independent of what we think, and we could be wrong. And, according to B2, we could be wrong about it 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, that I ought to comfort my partner. 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, 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. But they agree about the first. And it is the *conditionalisation* that I object to.

Of course, most theories allow for the possibility of moral error, so will conditionalise commitment 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.

For a *naturalist*, we may be entirely wrong about what matters if there is no external natural world, or if we are greatly mistaken about its contents. Certain kinds of judgement-dependence theorist conditionalise morality on what people want or will; if they become sceptics about other minds, they may then become moral sceptics. But these strike me as *perfectly good reasons* to suspend our commitment to morality. We all think that if there is no external world, or there are no other minds then nothing will matter morally (or what matters may be radically different).

On the other hand, 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. All our perceptual and empirical beliefs, all our inferences based on experience, could be *correct*, and naturalism still be true (or false). There might be no non-natural moral truths, and yet the rest of our beliefs about the 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 aspect of 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. All of these conditionals entail the Faithless Conditional, and can thus be seen to be objectionable. It would be a betrayal to our fellow agents if we abandoned our commitment to them just because we decided that the natural world is all there is. Even if we never made these discoveries - never concluded that God is dead, that the Mets have lost, or that naturalist is true - simply being prepared to abandon moral commitment should that situation come to pass, or thinking that there would be nothing wrong in so doing, is itself objectionable, is itself a kind of betrayal. We ought not to conditionalise our moral commitments on any of these things.

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 are 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 against non-naturalist realism. The rest of this paper focuses on objections to the argument I have just given.

One and the same?

My argument is that the non-naturalist must, objectionably, conditionalise one kind of thing - things mattering, or, more specifically, my 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 to whom I’m married doesn’t matter,” or “my partner doesn’t matter if no-one in my family matters.”

However, this response is question-begging, or, as Dasgupta says, not “playing fair”. (Dasgupta, MS) The realist makes substantive claims about non-natural truths, and what it takes to discover them, which don’t *obviously* apply to moral claims. IQ and Wisdom cannot be the same thing, since poor performance in a 30-minute multiple-choice test could not in principle be evidence for lack of Wisdom, but it is evidence for low IQ. Likewise, even someone who denies Naturalism will admit that parsimony considerations and questions about explanatoriness are at least in principle relevant to the question of whether naturalism is true (even though, of course, non-naturalists will deny that it is *conclusive* evidence). But I think 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 Wisdom *is* IQ would insist that a 30-minute multiple-choice test *can* be evidence for Wisdom, since it is evidence for IQ, just as 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 reader will agree that, on the face of it, the proposal that parsimony considerations are *in principle* not relevant to whether I ought to comfort my partner (or the claim that my partner’s mattering is not dependent on some non-causal, judgement-independent state of affairs), is just as plausible as the proposal that “I ought to comfort my partner” states a non-natural fact.

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 - 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. On Scanlon’s construal, there are certain “pure

normative facts” that are necessary (Scanlon 2014 p37). Likewise, Parfit claims that “The most fundamental normative facts...are necessary truths, which would be true in all possible worlds.” (Parfit 2011 II 307) According to this view, in any world which has the relevant kind of natural facts (eg, agents to serve as objects of moral concern), there will be something that matters. In all the cases where morality tells us not to be nihilistic, this theory tells us not to be nihilistic.

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 (and has) been realised - that the Mets lose - whereas the antecedent of Parfit’s conditional can never be realised. Such conditionalisations are trivially *true*, but no 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 who have exactly the same world-directed beliefs and attitudes as each other. To understand the nature of their disagreement, we need to look beyond the actual conclusions they draw about what conduct and attitudes are appropriate in this world; the Kantian objects to the Utilitarian not because the latter *actually* endorses lying, because he *would* endorse lying were things different, and indeed the Utilitarian thinks he *should* endorse lying in that different scenario.

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 have certain magical powers. So they are at least nomically 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 philosophical zombies indicates whether she properly appreciates the moral significance of phenomenal experience - zombies, lacking the latter, have a different moral status from normal people. But philosophical zombies are metaphysically impossible. Even 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, I accept that 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 following sense: 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. If moral thinking involves us planning for all the situations that we think we may find ourselves in whilst remaining rational, then 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 them negatively if they were to endorse, in this potential situation, that they then *ought* to become nihilistic, and abandon their moral commitments. But so long as they cleave to Ivan's and

Parfit's conditionals, respectively, then they do endorse nihilism in this situation. And that is what I object to.

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. I can 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. 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. 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.

Robust and Relaxed

Some will object that my arguments only tell against *one* kind of non-naturalist realist. "Robust" realists, like Enoch²⁵, accept substantive ontological commitments. But many prominent defenders of non-naturalism, such as Nagel, Scanlon and Parfit, have asserted that their claims carry no ontological weight - they are "relaxed" realists. Nagel says his realism should not be confused with an "inappropriate metaphysical picture" he refers to as "Platonism." (Nagel 1986 p139). Parfit says that reasons *exist* in a "wide" sense, and that this claim has "no positive ontological implications" (Parfit II 479) Asserting the existence of a non-natural, judgement-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

²⁵ As developed by his (2010). Bedke, in the quotation given above from his (2012) also accepts ontological implications, and Larmore endorses an ontological version of realism and suggests that Parfit ought to adopt it too: "Parfit asserts that truths about reasons 'have no ontological implications' (II, 486). But even on his narrow sense of 'ontological', this is not so." (Larmore 2013)

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.²⁶ Parfit thinks he is following Putnam (2004) in rejecting “Quine-inspired ontology.” But Putnam only allows truth-talk without ontological commitment because of his general pragmatism, which he draws from Dewey. And Dewey’s ethics is best seen as a form of naturalism or constructivism, so it is confusing that anyone might think that his views, however transmogrified, could support non-naturalism. In any case, Parfit does not endorse pragmatism - but Putnam’s rejection of Quinean ontology rests on it. Likewise, expressivist quasi-realists rescue truth from the jaws of ontology by appeal to a minimalist conception of truth. But neither Parfit nor Scanlon will have any truck with minimalism²⁷. So it feels like “relaxed” realism is a position defined purely negatively, by rejecting all the implications these philosophers dislike - leaving it quite unclear whether there is any coherent theoretical basis for doing so.²⁸

Perhaps the relaxed realist proposes, not a claim about ontology, but an expansion of *ideology*, in Quine’s sense. This makes sense of Scanlon’s view, where the talk is not of non-natural *objects*, but of non-natural *relations*. And this has some metaphysical appeal. It is not a form of Platonism in the sense Nagel wants to reject. In avoiding talk of a mysterious realm of strange entities, it is more comprehensible. It locates the truth-makers for moral claims *in* the world, albeit in an imperceptible web of non-causal relations that may or may not hold between things in the world. But it doesn’t avoid my objections. After all, the question of whether to expand our ideology, just like expanding ontology, incurs a theoretical cost. Expansions of ideology that don’t come with, say, explanatory

²⁶ As Larmore says of Parfit’s distinction between the “wide” and “ontological” sense of *exists* “What can Parfit have in mind with so perplexing a distinction?”

²⁷ “My view is not ‘minimalist’” (Scanlon 2014 p28)

²⁸ Parfit develops these views further in Volume III of *On What Matters*. However, I believe that the position presented there is confused, and so I find it more charitable to Parfit to reconstruct his position by drawing on the material in Volume II and similar statements made by his intellectual allies.

gain are thought to be prodigal, even prohibited. But these kinds of theoretical consideration are still irrelevant to the question of whether I should comfort my grieving partner.

Perhaps this is not how relaxed realists want to be understood. But *however* they 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, they must be entirely independent of what anyone happens to think - otherwise the view is not 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. And this is not just scepticism about other minds or the external world. I could be entirely correct about all the relevant natural and mental facts concerning my relationship with my partner - about our history, her feelings, the shared understandings that have bound us together - 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 scepticism which, 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 that is preposterous. On any available characterisation of the “non-natural facts”, learning of their non-existence could not be a good reason to abandon commitment to our partners.

Metaethics vs Ethics

My arguments against non-naturalist realism are *normative* 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 by pointing out what is implicit in familiar moral language - from the fact that moral claims appear to be cognitive, assertoric descriptions of states of affairs, and the fact that no naturalistic description seems like a sufficient basis for drawing a moralistic conclusion, and so on. Critics, as much as proponents, of non-naturalism have appealed to such descriptions of everyday moral thought and talk - Mackie argues for error theory by pointing out that ordinary moral talk appears to be committed to the existence of non-natural properties, and there are no such properties.

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. "Revolutionary" metaethics proposes that we *cease* speaking as though we were describing an order of non-natural facts, and start speaking as though we were expressing practical attitudes, or identifying solutions to practical problems and so on. It would be hard to claim that such a language could not *in principle* be used for the regulation of conduct - used, that is, as a moral language. 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 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 MS makes a similar claim). 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 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”, then he is marking out what he takes to be a basis for authority - fatherhood. And he can (and surely will) be criticised for this - 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? Why not take this as authority enough? 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 judgement-independent facts happens to support than we care about what we want, need and feel? To the extent that our morality is humanistic, then I think we can reject the argument from authority.

Dodging the Conditional

Can non-naturalism escape my challenges? Perhaps. 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 was

actually true, they would not become nihilists. They would simply adopt another metaethical theory. What basis do I have for criticising such philosophers? After all, even though they claim, *now*, that if naturalism were true then nothing would matter, they're not actually disposed to become nihilistic in that situation.

But it's not just enough to be *psychologically disposed* to have the right attitudes in counterfactual scenarios. My objection was not simply that non-naturalists *would* abandon their moral beliefs if naturalism were true. I've claimed that we *ought not* to abandon morality even if naturalism is true. That is a highly plausible first-order moral claim. Not to accept it would be a moral failing. It would be *missing* the fact that the truth of naturalism is not a good *reason* to change our attitudes. 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 naturalism is true.

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. If naturalism is true, they *should* become expressivists or naturalists or pragmatists or fictionalists. But they cannot, consistent with their own view, explain what moral grounds they have for doing so. If the non-naturalist is not simply someone whose psychology is well-described by Lewis' account, but thinks that they *ought* to carry on this way, then what possible grounds can they give for this claim? 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?²⁹

(By comparison, the Kantian's objection to the Utilitarian is not just that she *would* endorse lying if the utility-calculus showed it to be optimific - perhaps the Utilitarian in question is so

²⁹ Perhaps they believe that non-naturalism is self-effacing - that there are non-natural moral facts, but one of those facts is that no-one ought to care about whether there are such facts. If so, I have no quarrel with them. For my real goal is to argue for precisely that moral view - no-one should care if there are non-natural facts.

psychologically attached to conventional morality that she would, in reality, actually just drop her Utilitarianism in that scenario. The difference is that the Utilitarian, so long as she *is* a Utilitarian, at least thinks that she *should* endorse lying, whether she actually would or not.)

And there is another challenge. To endorse the Lewisian approach presumably requires one to think that some other metaethical theory could provide what really matters in ethics. All Lewis meant by calling such theories second-best is that they accord with our pre-philosophical assumptions about the ethical domain slightly less well, not that they fail to account for the normativity of ethics. Adopting, say, naturalism as a “second best” theory of ethics is not a matter of giving up on ethics and doing something else. It’s precisely that - adopting a different outlook on one’s own ethics. But if one thought that naturalism or expressivism *were* able to account for the normative authority of ethics, 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 those facts? The Lewisian approach robs non-naturalism of much of its motivation.

Indeed, adopting this view would rob non-naturalism of another of its signature features - explaining widespread moral error. According to the non-naturalist, some cultures and societies might be quite wrong in their ethics. They are wrong in that their views don’t correspond to the non-natural truths. It would be no vindication of such opposed views to point out that they actually track the extensions of those people’s moral concepts, or that they help to solve shared problems, or that they are just the kinds of principles that it would be useful to invent for people in their situation. But on what grounds can the non-naturalist then think *her* moral views could be vindicated by giving them a naturalistic, pragmatist or fictionalist reinterpretation? If non-correspondence with the non-natural condemns members of different ethical cultures, surely we are all thus condemned if there are no non-natural facts.

Conclusion

I want to close with a suggestion about what I think this shows regarding the relationship of ethics and metaethics. After all, nothing I have said does anything to show that there *are* no 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. Clarke-Doane has suggested (for different reasons) that even knowledge of the *normative* facts might not be enough to settle what to do. I think that's right, because one would have first needed to determine that the normative truths were worth caring about. 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 or demonstration that would compel even the recalcitrant amoralist to be moral was "the holy grail" of Ethics.³⁰ I think most philosophers 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 (wrongly, in my view) 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

³⁰ 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 be hostage to such a search. We don't altogether approve of Holy Grails. We do not see the need for them." (Blackburn 2010 pp127-128)

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.

Bibliography

Bedke, Matthew (2012) "Against Normative Naturalism" in *Australasian Journal of Philosophy*, 90:1, 111-129

Benacerraf, Paul (1973) "Mathematical Truth." *Journal of Philosophy* Vol. 60: 661–679

Blackburn, Simon (2010) "Must We Weep for Sentimentalism?" in *Practical Tortoise Raising and Other Philosophical Essays* (Oxford, Oxford University Press)

Coons, Christian (2010) "How to Prove That Some Acts Are Wrong (Without Using Substantive Moral Premises)" in *Philosophical Studies* 2010; 155: 83.

Clarke-Doane, Justin (2012) "Morality and Mathematics: The Evolutionary Challenge," *Ethics*. Vol. 122. 313-340.

Clarke-Doane (2016) "Debunking and Dispensability" in Neil Sinclair and Uri Leibowitz (eds.), *Explanation in Ethics and Mathematics* (Oxford: Oxford University Press.)

Dasgupta, Shamik (MS) "Non-Naturalism and Normative Authority"

Dworkin, Ronald (1996) "Objectivity and truth: You'd better believe it" in *Philosophy and Public Affairs* 25 (2):87–139 (1996)

Enoch, David (2010) "The epistemological challenge to metanormative realism" *Philosophical Studies* 148 (3):413 - 438 (2010)

Enoch, David (2011) *Taking Morality Seriously: A Defense of Robust Realism* (Oxford University Press)

Dummett, Michael (1978) *Truth and Other Enigmas* (Cambridge MA: Harvard University Press)

Field, Hartry (1989) *Realism, mathematics, and modality* (Oxford: Basil Blackwell)

James, William (1956) "The Moral Philosopher and the Moral Life" in *The Will To Believe* (New York)

Larmore, Charles (2013) "Morals and Metaphysics" in *European Journal of Philosophy*, 21: 665–675

Lewis, David (2005) "Quasi-Realism is Fictionalism" in Mark Eli Kalderon (ed.), *Fictionalism in Metaphysics* (Oxford University Press)

Nagel, Thomas (1986) *The View From Nowhere* (Oxford University Press)

Parfit, Derek (2011) *On What Matters, Volumes I & II* (Oxford: Oxford University Press)

Railton, Peter (1984) "Alienation, Consequentialism, and the Demands of Morality" in *Philosophy and Public Affairs*, Vol. 13, No. 2. (Spring, 1984), pp. 134-171

Scanlon, T.M. (2014) *Being Realistic About Reasons* (Oxford University Press)

Street, Sharon (2006) "A Darwinian Dilemma for Realist Theories of Value" *Philosophical Studies* (2006) 127:109166

Väyrynen, Pekka (MS) Normative Commitments in Metanormative Theory

Williams, Bernard (1973) "A Critique of Utilitarianism," in *Utilitarianism For and Against* (Cambridge: Cambridge University Press)

Williams, Bernard (2006) "Philosophy as a Humanistic Discipline in *Philosophy as a Humanistic Discipline* (Princeton University Press)

Paper Four

Sentimentalist Pragmatism Defended

For some philosophers, to see ethics as invented rather than discovered, as derived from our sentiments rather than from reason, as highly contingent in its content, as anything less than fully universal and objective, would be to undermine its authority, significance, or legitimacy. It may in fact be that the practices that we designate as ethical are just like this - that they're merely the sorts of things that creatures like us with feelings like ours invent in order to solve their own problems. But to assert such a thing would be akin to asserting moral scepticism or nihilism. If that's all there is to ethics, then the pretension of ethical prescriptions to govern our conduct is a sham. If there are no objective, mind-independent, universal, rationally compelling ethical facts, then ethics is normatively groundless.

I think that ethics really is invented. Ethical principles and practices are things we created to solve the kinds of problems we find in living together. And by "we" I mean something quite limited. Perhaps there are practical problems that all rational creatures would find themselves to have. But it does not matter if there are not, or if these are not the kinds of problems to which ethics as we know it offers solutions. I mean creatures with a certain subset of pro-social sympathies. Entirely unsympathetic creatures would find entirely different situations problematic, and would invent different solutions to them. And creatures whose capacities for sympathy extended only as far as their tribe, swarm or kin-group would take themselves to have different problems again. But for creatures like us, whose capacity for mutual concern can be extended far beyond the bounds of biological affinity, the kinds of thought, practice and speech that we call "ethics" are valuable, because it solves our problems. As our societies develop and change, so too will those problems, and with it the contents of our ethical outlook. Ethical life involves ongoing inquiry, not because there

are objective ethical truths that lie beyond our ken, but simply because living together in a changing world requires constant thought and experimentation to see what works.

I think these claims about the nature and origins of ethics are true as a matter of fact. But my concern here is not to establish those descriptive claims. Rather, I aim to meet the charge of moral scepticism or nihilism raised at the start. Many philosophers have rejected views like mine - those which are sentimentalist, constructivist, naturalistic or pragmatic - on the grounds that they miss out a vital ingredient in explaining the authority or normativity of ethics. This itself, in my view, is a normative claim. To say that moral claims would lack authority if they had a certain origin or nature is to deny them normative, reason-giving, action-guiding force. And to deny that a claim has action-guiding force is just as much a normative judgement as the contradictory assertion. To say that a purely naturalistic, mind-dependent, emotion-driven, contingent ethics wouldn't "matter", to use Parfit's famous terminology, is itself a claim about what *does* matter. And so my concern here is normative. I think that if ethics has just the features that I attribute to it, it would still matter, it would have authority, it would be exactly as legitimate as committed ethical agents could hope. If contingent human sentiments and variable human problems are all there is to ethics, that is quite enough to explain why we should care about ethics.

In what follows, my strategy is to consider a series of putative desiderata that some philosophers suppose ethics needs to satisfy in order to claim practical authority - mind-independent truth, objectivity, a basis in rationality, and universality. I ask whether these really are things that we should regard as indispensable. My argument is that they are not. We can explain why we revise and refine our ethical thoughts in the light of intelligent inquiry without appeal to objective truth, and we can explain the nature and extent of interpersonal moral criticism without supposing that ethics has a universal basis.

My goal is not to provide a thoroughgoing argument in favour of my own metaethical view, which I call sentimental pragmatism. I recognise that my portrayal remains underspecified - without a more detailed account of what makes something a practical problem, what forms of sympathy and pro-social sentiment I appeal to, and a fuller account of the nature of moral reasoning, this view cannot be fully assessed. And I do not aim here to explain why I diverge from other views that are broadly similar, such as Kitcher's ethical pragmatism (2011), which places a greater weight on the notion of progress than I do, or Lenman's expressivist constructivism (Lenman 2012), which lacks my emphasis on the notion of practical problems.

I only argue here that an account of this form can ground the authority of ethics, and that features of other prominent metaethical theories are at least not needed for this. But just because mind-independent truth, objectivity and universality are not *needed*, that does not mean that they cannot be found. For all I say here, perhaps ethics is like this. But I have argued elsewhere that strong forms of realism are actually unacceptable - that they undermine the normative authority of ethics (Hayward MS *b*), and undermine our practices of moral inquiry (Hayward MS *a*). Taking these essays together, my hope is to have provided a strong motivation for exploring theories that dispense with mind-independence, objectivity, and universality.

Methodology: between ethics and metaethics

Questions about truth-aptness, objectivity, mind-dependence and realism in ethics are often regarded as belonging to *metaethics*, as opposed to normative ethics. Normative ethics asks questions about what is actually good and bad, right and wrong, valuable and valueless. According to many philosophers, such as Mackie (Mackie 1977 Ch1), the more abstract questions of metaethics are entirely neutral with regard to normative ethics. Metaethical controversies leave normative ethical claims where they stand.

Concomitantly, it's often thought that metaethical questions cannot be established or argued for on normative grounds - metaethics is a theoretical, not a practical discipline, and metaethical claims are to be established by linguistic, psychological, metaphysical or other theoretical arguments. As Blackburn has suggested, the role of the metaethicists is that of an "anatomist" or "anthropologist," who simply observes and describes the features of moral life, rather than engaging in ethics or prescribing normative views in the manner of the "moralist". (Blackburn 2010)

So there are two claims here - first, that metaethical theories do not affect normative ethics, and second, that normative ethical claims are not legitimate bases for selecting our metaethical theories.

But the first claim is surely implausible. Normative ethics is not a finished discipline; rather, all of us take our normative viewpoints to be open to revision in the light of further inquiry. As such, we accept certain standards of evidence as reasons to change our normative beliefs. And our standards of evidence are dependent upon what we take truth or correctness in ethics to consist in.

For example, our assessment of the probative value of moral intuitions depends on our views about the relationship between moral correctness, rationality and the emotions - rationalists will find the sorts of arguments presented in Singer and Greene (Singer 2005; Greene 2008) against deontological theories persuasive in a way that sentimentalists will not. (Singer 2005; Greene 2008) Likewise, evolutionary debunking arguments such as those presented Street and Joyce are explicitly designed to work only against metaethical theories that take moral truth to be robustly mind-independent; ethicists who deny this kind of independence needn't fret about the evolutionary genealogy of our moral views. (Street 2006; Joyce 2006) "Local" debunking arguments, which attempt to undermine only particular moral ideas or styles of thought, are, as Kahane has argued just as dependent on metaethical assumptions about what correctness in ethics consists in as the global debunking of Street and Joyce. (Kahane 2011) As such, I am in agreement with Bloomfield

who argues that moral epistemology provides a bridge between metaethics and normative ethics - our views about moral truth and correctness shape our standards of evidence, which in turn shape our substantive first-order theories. (Bloomfield 2009)

If metaethics affects normative ethics, is it really plausible that no normative assumptions or claims are relevant in selecting our metaethical theories? To be sure, there are some aspects of metaethical inquiry where Blackburn's portrayal of the anatomist is apt. Sometimes, our goal in metaethics is simply to describe how those who engage in ethics happen to speak or think, or what metaphysical assumptions they happen to make. To the extent that this kind of research is purely descriptive, it is plausible to think that normative ethical claims are irrelevant to establishing the accuracy of any metaethical theory, just as they would be in establishing the truth of some natural- or social-scientific description.³¹

But, of course, if *that* were all metaethical argument amounted to, the sorts of moral-epistemic and thence normative conclusions that I just described could not follow from such metaethical theorising. "Anatomical" metaethics just tells us how people happen to think, speak and feel when they engage in ethics. And, as such, it may well give us an account of what "the folk" take to constitute truth or correctness in ethics. But that is not the same as actually arguing that we *ought* to accept any given account of ethical truth or correctness. That would be an is-ought inference of the crudest, and most uncontroversially unacceptable, kind. In order to constrain our moral epistemology and hence our normative ethical viewpoint, metaethics needs to tell us what account of ethical correctness we actually should adopt.

This cannot be a purely descriptive matter, because even if we do happen to think and speak in just the ways that one metaethical theory claims, we always retain the option to do otherwise. As

³¹ Of course, as an epistemic pragmatist I *do* think that ethical considerations are relevant for establishing the *justification for believing*, but this is beside the point.

“revolutionary” metaethicists propose, we could always change our practice.³² And understanding the normative implications of our account of ethical correctness might give us the impetus to do just that. For example, even if non-naturalist moral realism is actually a correct account of folk moral practice and thought, we might, in the face of evolutionary debunking arguments, conclude that we should deny that non-natural truth-makers are necessary to give authority to morality, instead adopting a different account of moral correctness, one that is not vulnerable to debunking arguments.³³ To the extent that our goal in metaethics is not simply to describe folk practice, but to actually establish some account of ethical truth or correctness as the one which we ought to adopt, purely descriptive arguments will not do. As Väyrynen (MS) has argued, normative assumptions are indispensable in establishing such metaethical conclusions.

Some philosophers propose that metaethical claims about the nature of ethical truth or correctness can be established simply as conceptual or constitutive truths (eg Cuneo & Shafer Landau 2014 for), without having to appeal to substantively normative arguments. But such arguments are either implausible, or they are simply disguised attempts to move from anatomical “is”es to normative “oughts”. It may be true of certain *conceptions* of morality that moral truth *must* be objective or mind-independent. But so what? This is just an anatomical claim. It may also be true of certain conceptions of marriage that marriage *must* be between a man and a woman. But other conceptions of marriage are available, and egalitarians can choose to adopt them.

So the conceptual argument would need to assert that no other conceptions of morality are possible, or could play the role that the concept of morality is supposed to play in our thought and practice. But this is surely implausible - why not think that a fictionalist, constructivist or expressivist language, with its attendant conception of morality, could at least in principle be used to regulate

³² The notion of “revolutionary” metaethics is a generalisation of the notion of “revolutionary fictionalism” in metaethics advanced by philosophers such as Joyce (eg Joyce 2016). The “revolutionary” vs “hermeneutic” distinction was originally introduced into the nominalism debate by Burgess (1983).

³³ This is the argument of Hayward (MS *b*); I also read Street 2006) as saying something similar.

conduct, offer criticism, and make appeals to other agents? The objections to these anti-realist views is not that such conceptions of morality could not be used to regulate conduct, but that they lack something that morality needs if we are to grant it authority. That is to say, they are normative objections.

Likewise, it may be constitutive of a certain *kind* of agency that it seeks to conform itself to universal rational principles. But it takes a normative argument to establish that we *ought* to concern ourselves with *that* kind of agency, rather than contenting ourselves with some more minimal kind of “schmagency” (Enoch 2006). More could be said here, but I remain unconvinced that any metaethical theory established on conceptual or constitutive grounds, without implicit appeal to substantive normative standards, could constrain our moral epistemology, and hence our metaethics, in the manner I described above.

As such, the sort of metaethics that I am interested in here starts not from the question “what are the fundamental features of (folk) moral thought”, but rather such questions as “What would be needed for ethics to matter, to claim legitimate authority over us?” or alternatively, “What are the features such that, if ethics had it, ethics would have a claim to be genuinely action-guiding.” Like any questions about authority or legitimacy, these are normative questions. These could not be “external” questions, in Carnap’s sense - questions that we could sensibly ask and answer whilst standing outside of and aloof from any particular normative perspective. (Carnap 1950) Rather, this investigation will be “internal” - a piece of substantive normative theorising, that can only be undertaken by those who already accept certain normative principles or judgements. If there is disagreement about what ethics needs to be like to have authority, it will be disagreement about what matters and is worth caring about, not simply disagreement about metaphysics or language.

I am not the first philosopher to propose that we select our metaethical views on normative ethical grounds. Dworkin (1996) famously argued that those who denied objectivity and mind-independent truth to ethics were doing something substantively immoral (Kramer (2009) has made similar arguments). And, indeed, I think that many of the philosophers who have accepted realist, objectivist, universalist or rationalist views in metaethics have done so because they believe that only such accounts can underwrite the legitimacy or authority of ethics. I agree with Dworkin and his followers that we need to ask what ethics needs to be like - or, perhaps, what the world needs to be like - for ethical claims to have authority, and I agree that this is a substantive normative question.

Where I differ is in suggesting that mind-independent truth, objectivity, rational obligatoriness, universality and all the rest are really necessary for ethics to be authoritative. Even if ethics is invented, contingent, emotion driven, and local, it is still authoritative for creatures like us. Or so I shall argue.

What is Authority?

Talk of the authority of morality (or of normativity, or of reasons) is so common that it is easy to avoid ever having to spell out what precisely this means. It is often critics of views like mine who talk in terms of authority, and I am not always entirely sure that they have a clear conception in mind. For example, Joyce says that since moral claims are backed up with non-natural, purely normative truthmakers, they lack “practical clout” or “oomph” (Joyce 2006 p8). This obviously calls out for more detail.

Dasgupta (MS) distinguishes between two senses of authority in contemporary metaethics. One is a psychological conception, concerning the link between normative judgement and motivation, which, as he says, has exercised more interest among metaethicists. What is the psychological mechanism which brings us to be motivated to act in accordance with morality’s commands? Is it

the case that we are always motivated to act in accordance with our moral views, and if so, what explains this? Why, in other words, do we care about morality? Of course, the psychological question is not new - for example, Mill is quite clearly concerning himself with the same issue when he talks of the “sanction” of the principle of utility (Mill 1979 Ch3).

The question of morality’s psychological authority is certainly an interesting one, and an important topic for descriptive moral psychology. But it is not my topic here. When I say that morality has authority, I don’t just mean that people are, in fact, motivated to be moral (or, since without qualification *that* is clearly false, that some people are sometimes motivated to be moral). When I ask about the authority of morality, I am asking why we *ought* to do what morally tells us to do, not why we want to do so. That’s why I said that the question of authority is a normative question, and one that can only be answered on normative grounds. It is a question *internal*, in Carnap’s sense, to normative ethics. To say that morality has authority in the normative, rather than psychological, sense, is to say that sometimes, people ought to do what morality commands - even when they wouldn’t be otherwise inclined to do so.

This final qualification - the idea of there being something you ought to do even if you wouldn’t otherwise want to - can lead some philosophers to conflate *authority* with *categoricity*. Whereas *hypothetical* imperatives are claims about what you ought to do given certain goals or desires you have, *categorical* imperatives are claims about what you ought to do that make no such reference to any of your goals. But notice that the question of authority arises even in the case of *hypothetical* imperatives - often, agents have no inclination at all to take up the means to their ends. To say that they *ought* to do so is to attribute authority to the norm of instrumental rationality. And it is far from obvious that this norm has the authority sometimes attributed to it.³⁴

³⁴ For an argument against the authority of the instrumental norm, see Hayward 2017

So my view is that philosophers have made a mistake in focusing on categoricity as the key to explaining the special status of moral normativity. The difference between hypothetical and categorical imperatives is simply that the latter are supposed to have a broader audience than the former. But they are alike in that both are imperatives or prescriptions. For any given prescription, one might always be disinclined to obey, and one can always ask “why should I do that?”. To explain why someone ought to follow a prescription is to explain its authority, and to do that is to make a normative claim.

Truth and pragmatic correctness

The notion of correctness plays a crucial role in the vindication of moral authority. There is a myriad variety of moral perspectives, sets of norms, and theories, both possible and actual. And yet there is only one set of moral claims that I take to govern me and my actions - the set of claims that I actually accept. But if I thought that this set of claims was no more correct than any others, then why respect its prescriptions?

One point of difference between my moral outlook and others is precisely that it is *mine*. But, judged from the perspective of an ethical agent, the mere fact that a moral claim is one I currently believe does not seem to grant it authority over me. After all, as I mentioned before, no-one but the most close-minded dogmatist would claim that their current moral outlook constituted the endpoint of all moral inquiry. Rather, we take our own ethical outlook to be, at least in principle, revisable in the face of future inquiry. Such revisions, we take it, may stand to *improve* our moral views. To make sense of this fact, we need some conception of a standard of moral correctness that is distinct from the list of individual moral beliefs and principles that we happen to adopt currently. Without such a conception, the revisions wrought by inquiry seem nothing but “mere change” (cf Kitcher 2011 p138) - no better and no worse than what went before.

If, as I think, there are no mind-independent moral facts or truths, how can we make sense of these twin phenomena - the claim that some ethical views are correct (and others incorrect), and the idea that there is a standard of moral correctness by reference to which revisionary inquiry can lead to improvements? If ethics is invented by humans, as I believe, then how can it answer to any such standard of correctness? For many philosophers, it has seemed obvious that the reason that I should obey *this* set of moral prescriptions rather than any available alternatives is because it is closest to the *truth*, robustly construed; moral inquiry improves our moral beliefs by bringing them into closer correspondence with the moral *facts*.³⁵

I agree that we need a notion of correctness in order to vindicate the authority of morality and make sense of the value of inquiry. But the notion of correctness, of a standard by which we could measure change as better or worse, is broader than the notion of correspondence truth. A *pragmatic* notion of correctness in ethics defines correctness in terms of *practical problems*. Ethical pragmatists deny that our goal in ethics is to have beliefs that correspond with purely ethical facts or truths. Rather, ethical agents find themselves with certain problems in acting, and ethical judgements help mark out the solutions to those problems. By following a particular prescription, rule or principle, adopting a certain goal, valuing a particular property of state of affairs, acknowledging some standard of virtue or excellence, ethical agents, individually or jointly, contribute to the solution of the problems they face. Normative judgements about *which* norms, principles, standards of virtue, values or rules these agents *ought* to adopt are correct to the extent that they indeed demarcate successful solutions to the problems in question.

The notion that ethical correctness can be defined in relation to practical problems is sometimes taken as the defining feature of *constructivist* views in ethics (Korsgaard 2003, Street 2008). However, at other times, constructivism is taken to be the narrower, more exacting claim that ethical

³⁵ Michael Smith's voices a typical view "It is distinctive of moral practice that we are concerned to get the answers to moral questions *right*" (Smith 1991)

correctness is defined as the output of a *rational decision procedure* (James A. 2012), or, more broadly, as any view on which values are a product of the evaluative attitudes of agents, without any reference to the specific notion of a practical problem (Silk 2015). As such, I will leave this much-contested term “constructivism” to one side, and use “pragmatism” to pick out any view that defines ethical correctness in terms of problem-solving.

A pragmatic standard of correctness can make sense of the claim that some ethical views are better than others; it explains how revisionary inquiry can lead to improvements rather than “mere change”. After all, relative to a given problem, a set of moral ideas, principles and standards can do a better or worse job of providing solutions. So for the pragmatist, some ethical views really are better than others. And the notion of correctness outruns what we currently happen to accept. It may be that our current ethical views do not actually solve the problems that they are meant to solve, or that alternative views would solve our problems better, or that new problems arise which call for new and novel solutions. Relative to the standard, there is a clear sense in which the products of revisionary inquiry can constitute an improvement on what went before.

I have argued that we can make sense of the idea that there are standards of correctness and improvement in ethics without appealing to any robust conception of ethical truth. Without a standard of correctness, I agree, ethics does seem to lack normative authority. But it does not follow that the standard I have described is sufficient to underwrite that authority. For there is an obvious lacuna in the pragmatist position offered so far; or, to put it another way, pragmatism marks out a style of metaethical theory, but no metaethical theory in particular. What practical problems does morality attempt to solve? Or, since our interest is normative rather than descriptive, is there any way of identifying these practical problems that preserves the idea that morality, in solving them, is something that really matters, that legitimately claims to govern us?

Kinds of Objectivity

My theory differs from other pragmatic theories in its specification of the practical problems that ethics should seek to solve. Before I turn my attention to these other accounts, I first want to address an accusation that all theories of this sort face. For a common complaint against such views is that they cannot account for the *objectivity* of ethics. And if ethics is not objective, goes the complaint, then how could it possibly be authoritative?

There is one sense in which pragmatic theories clearly do make evaluations of moral correctness an objective matter. Given a particular problem, it will be an objective fact whether some moral proposal really constitutes a solution. A tribe or society may struggle to live together harmoniously, and adopt shared norms and standards of virtue to regulate their interactions. If the norms they have adopted lock them into an endless cycle of honour-killings, vendettas, and strife, then we can say with perfect objectivity that their moral outlook has not solved their problem.

But we may worry that a *deeper* sense of objectivity is needed. Even if it is an objective fact, relative to a problem or set of problems, whether some moral outlook provides solutions, doesn't there need to be an objective basis for determining *which* problems are the ethical ones? Maybe my hypothetical tribe should not take vendettas and strife as a problem, or harmony as an ideal state. After all, just as there is a myriad of potential ethical outlooks we could adopt, there is equally a huge array of issues that we could identify as key problems in needs of ethical solution. I said that it is a *normative* question what problems ethics should seek to solve. But whose norms count here? Isn't this a point where we need to appeal to objective normative truths?³⁶

Although this is a common thought, it's hard to know what this deeper sense of objectivity amounts to. Sometimes, talk of the objectivity of morality is elliptical for the claim that moral prescriptions

³⁶ Hussain and Shah make this point against Korsgaard in their (2006).

are binding on people whether they want to be moral or not. But this is simply a restatement of the claim that morality is authoritative, not a further condition of what morality has to be like in order to count as authoritative. My view is that human beings really ought to be moral, whether they want to or not, regardless of whether morality is objective in any deeper sense. So, again, I think it is a perfectly objective matter, in this sense, that we ought to take the problems we encounter in living together as our basis for moral inquiry and action. *You* should care about those problems, whether or not you would otherwise be inclined to do so.

But again, this might seem like an insufficient response. Even if I can convince you that sympathetic interpersonal problems are the sort of things you ought to care about, and that you would do well if you assessed the correctness of moral conceptions by reference to their success in solving these problems, the fear might remain that this is “just us”. Perhaps our agreement here simply reflects our human psychological quirks, or worse, our evaluative prejudices. We agree that these problems are worth caring about, and that their solutions ought to be granted a certain authority, only because we happen to be sympathetic, social creatures. But perhaps we are wrong? Perhaps this is not, ultimately, what is valuable?

If the worry is that the particularity of our viewpoint has clouded or biased our assessment of what problems are worth solving, then perhaps the sense of objectivity that is needed is a perspectival one. Williams spoke of an “absolute conception” of the world (Williams 1978 p64), “to the largest possible extent independent of the local perspectives or idiosyncrasies of enquirers, a perspective of the world...’as it is anyway” (Williams 2006 p184). This is similar to Nagel’s conception of objectivity, which is formed by progressively retreating from the limitations and particularities of our own subjective perspective in order to seek a “view from nowhere”. (1986) I am not sure that such a conception makes sense, and my own views on the nature of the ethical perspective do not depend

on any contrast with such an absolute point of view. Still, perhaps this is what philosophers have in mind when they worry whether their normative views are objective?

On this conception of objectivity, then, a problem is objectively normatively significant if a creature utterly different from human beings - as different as it would be possible to be whilst still being capable of any kind of thought and perception - would still take the same problems to be worth caring about. Williams thought that the absolute conception included only the claims of physics - a radically different kind of cogniser would be bound to appreciate physical facts, but could quite adequately navigate the world without any sensitivity to the moral or the aesthetic. Nagel, though, found it quite plausible that a perspective abstracted from human particularity, and in particular from human sympathy and sentiment, would still be one from which the same moral issues appear to be normatively significant.

Is it plausible to say that morality is only authoritative if a detached, emotionless, unsympathetic intelligence would take moral problems to be worth caring about? Do moral problems have to seem significant when seen “from nowhere” in order to have authority over us? I think not. Why care about what such a miserable creature would think? Williams bridled against those who painted his theory of the “absolute conception” as crude scientism (Williams 2006). To say that physics, and not morality or aesthetics, might come closest to the view of the world that would be seen from a perspective maximally cleansed of human particularities would not thereby be to denigrate the latter two fields. Why think it any great mark of honour to be visible from this inhuman perspective?

Perhaps moral problems really do seem significant viewed from the absolute perspective, the view from nowhere, or the objective standpoint. I do not know, for I have never occupied that standpoint. But if morality is something only experienced by those equipped with the full range of human concerns, whose hearts beat responsive to one another (James 1956 p196) and whose minds have the

power to gild and stain the world and raise it in new creation (Hume 1994 p294), then it strikes me as a kind of contempt for humanity and for the human perspective to think any the less of it on that score.

Internalism and criticism

But if the problems that I designate as moral only appear significant from a particular perspective, then what are we to say to those agents who do not share that perspective? Such agents, after all, at least seem possible. There are two worries here. The first is that denying that morality exhibits the kind of perspectival objectivity I mentioned in the previous section implies a form of relativism. The second is that this form of relativism is incompatible with viewing morality as authoritative.

On my view, ethical correctness is defined in relation to a certain set of practical problems, and only agents with particular, contingent features of their psychology will find these problems to *be* problems - that is, to be normatively significant, and worthy of solution. It is possible that there are agents who are quite indifferent to the concerns that I designate as lying at the heart of ethics. So there is a legitimate sense in which this view could be called relativism - ethical correctness is defined relative to the perspective of sympathetic, social creatures.

But we should distinguish this *perspectival* conception of relativism from the *normative* claim that agents who lack moral motivations are somehow beyond the boundaries of ethical criticism. After all, the latter claim is not obviously entailed by the former - just because caring about morality requires one to occupy a certain perspective, it doesn't immediately follow that those who do not occupy that position are indemnified against moral criticism. And if any form of relativism is a threat to the authority of morality, surely it is the normative form. Without further argument, the mere fact that moral correctness is defined relative to the perspective of a sympathetic, social

creature seems no more of a threat to morality's authority than the fact, if it were a fact, that moral correctness is defined in relation to the absolute perspective.

It might be, though, that perspectival relativism actually does entail normative relativism, and that normative relativism does, indeed, undermine moral authority. Here is an argument to that effect. Many philosophers accept some version of Kant's dictum that "ought implies can". There are, of course, a huge variety of interpretations of this claim, and certain scholars who deny it altogether (eg Buckwalter & Turri 2015; Chituc, Henne, Sinnott-Armstrong & De Brigard 2016). But according to some philosophers, such as Griffin (2015), the "can" in question includes *motivational* abilities. If an agent *cannot* be motivated to perform some action, it follows that it is not the case that she ought to do this. Even if the agent could physically perform the actions required by morality, if she is unable to acquire distinctively moral motivations, to occupy the perspective of one who finds morality to be motivating or significant, then it follows that she is under no obligation to respect morality's prescriptions.

This is why it matters if morality is absent from the objective perspective; if it were, then we might say of any agent that exhibited a reasonable degree of cognitive sophistication *could* step back, take a more objective stance, and gain moral motivations. If, as I claim, moral motivation stems from sympathy and other pro-social sentiments, then it may well be that there are agents who simply cannot occupy the moral perspective, agents that are indifferent to one another, callous, or even irremediably cruel. Of course, not all wicked agents are entirely *incapable* of occupying the moral perspective. Most could be more sympathetic, benevolent and so on, and simply fail to do so. But perhaps some of the worst villains in human history were like this - psychologically incapable of moral motivation. We are tempted to say that these agents behave morally wickedly, that they ought to do better. Yet according to the argument given, such agents would be beyond the scope of moral criticism.

Suppose we accepted this argument. Why might this be thought to undermine the authority of morality? After all, even if there is a motivational constraint on moral obligation, at least *we* - that is to say, sympathetic, social agents - are capable of being motivated to solve moral problems. So nothing stops us from having moral obligations. In what sense, then, might perspectival relativism undermine the authority of morality? A plausible answer: it is not enough for morality to have authority only over the already somewhat decent - the sympathetic and the social - it must provide a basis for criticising the utterly cruel, indifferent, and callous. If ethics does not give us a basis for criticising the very worst agents, then that does seem like a deficiency of authority.

So is the argument indeed correct? I think that it rests on a confusion. This confusion is often found, I think, in contemporary interpretations of Bernard Williams, whose famous “internalism about reasons” is perhaps the most influential source for the argument that (some forms of) moral obligation are dependent upon the motivations of agents. As Williams claims, agents who could not come to be motivated to be moral, via a sound deliberative route from their current motivational set, cannot be said to have a *reason* to obey morality’s prescriptions. The man who has no motivation that would be served by ceasing to beat his wife cannot, Williams claims, be said to have a reason so to cease. (Williams 1981) Interpreters sometimes seem to suppose that Williams is thus asserting that they are beyond the scope of moral scrutiny altogether.

But this overlooks the fact that Williams meant something quite specific when he denied the wife-beater had a *reason* to desist from violence. In the current philosophical scene, talk of reasons is often seen as fundamental to the normative domain, such that all moral claims must be ultimately reducible to claims about reasons. On this view, if we cannot say that an agent had a reason to behave other than he did, there is no other weighty normative criticism that we can make of him. But Williams attributed a much more limited role to reasons-talk. This is why he insisted that, while

the brutish wife-beater may have no “reason” to refrain from his abuse, he may certainly be condemned as cruel, wicked, beastly and all the rest. To Williams, these criticisms were just as weighty - if not more so - than claims about reasons.

My view is that there is an important insight in Williams’ position, but it does not lend support to the argument I have been considering. To understand Williams’ view, we need to realise that moral criticism and condemnation are not monolithic. Sometimes, the point of moral criticism is simply to condemn, to mark out the wicked as the degenerates they are, and to marshal the forces of virtue against them. But at other times, criticism has a more constructive purpose - we use moral criticism to engage with others, to persuade them to do better. The internalist is right if he tells us that criticism of the second kind is misguided when we deal with agents who simply cannot be motivated to behave morally. This is true for no deeper reason than that it would be a waste of time. The brutish wife-beater cannot be persuaded to reform by reasoning with him, so it would be foolish to try.

But it does not follow from this fact that criticism of the first sort is ruled out. After all, to say that we should not criticise the callous and the cruel would itself be a normative claim, and one that carries little persuasive force. Nothing in perspectival relativism tells us to refrain from condemning the wife-beater as the brute that he is, from marshalling our forces against him and hating him for his depravity. And if we can do these things, then I would say that the authority of morality has not been seriously undermined. To acknowledge that some agents cannot be reasoned with, to recognise that they are incorrigibly immoral, is not thereby to give them a free pass.

In fact, the distinction between these two forms of criticism is implicit in everyday moral talk. When an agent is considered entirely beyond the pale, as irremediably deficient in his moral psychology, then moral observers will frequently talk about how wicked he is, or how important it is to stop him,

but rarely about what he ought to do, or what he has reasons to do. Many progressives hold precisely that view of Donald Trump - he is so egoistical and unprincipled that he simply could not be brought to share the concerns of his liberal critics. And thus it is rare to find commentaries from those same critics arguing that Trump *ought* to do this or that, or that he has a *reason* to enact more tolerant or socially just policies. Rather, focus is on organising resistance against him and his agenda, or on diagnosing and skewering his perceived vices. In fact, the only political journalism that I have read which talks about what Trump “ought to” do is written by those who are at least partially sympathetic to his political orientation.

So my view is that condemnation can continue even when persuasion is powerless. The fact that some agents cannot and will not be motivated to care about moral problems does not undermine the authority of morality in any way worth worrying about. The assumption that morality would lack authority if we could not, in addition to condemning the wicked, also reason them into better behaviour, arises, as Blackburn writes, from:

...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... [This wish] 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 be hostage to such a search. We don't altogether approve of Holy Grails. We do not see the need for them. (Blackburn 2010 pp127-128)

It would certainly be nice if we could find some way to persuade even the most recalcitrant immoralist (or amoralist) to share the concerns of decent moral agents. It would make the pursuit of moral goals easier, and the power of moral arguments greater. But if there is nothing that can do this, then we should not fear for the authority of morality. We who are suitably motivated can still get on with the practice of deliberating about what to do, debating with each other about what norms to endorse, and enforcing and policing compliance of those same norms among any who live in community with us.

Rationality and reasoning

Nevertheless, there is a legitimate worry in the view that Blackburn is criticising. If too many of the world's knaves lay beyond the scope of reasoned refutation, then that would indeed constitute a threat to authority - not the authority of moral norms, but of moral reasoning. After all, ethics is not just a set of rules, evaluations and prescriptions, but also a characteristic pattern of thought and deliberation. Participants in the ethical life give reasons for their judgements, and work on the assumption that these reasons have some persuasive power in bringing others into the circle of moral commitment. Some philosophers have rejected sentimentalist views precisely because, they claim, they do not give sufficient significance to the power of moral reasoning.

Even if there is nothing to stop us from simply condemning those that cannot be persuaded to be moral through reasoning, we should worry if we are forced to resort to this option too often. If the standard of correctness in morality was indexed to practical problems that only a select few could be concerned with, then our practices of moral reasoning would be undermined. They would lack the significance we often give to them. But this is clearly a matter of degree. It is surely too much to ask that moral reasoning be capable of persuading *all* agents. Everyone ought to accept that sometimes persuasion gives out, and we have to resort simply to condemning and confining the wicked. So the question is - how expansive does the potential audience of ethical reasoning need to be, and how extensive does the power of persuasion have to be, in order to make sense of the practices of ethical deliberation that we are engaged in? Does my view, that morality is something that only those who are capable of certain kinds of pro-social sentiment are capable of participating in, limit the power of moral reasoning too far?

Kantian constructivists, like Korsgaard, agree with me that correctness in ethics is a matter of solving practical problems, rather than corresponding to mind-independent moral facts. As such, Kantian constructivists are generally classified as moral anti-realists, like me. They also deny

perspectival objectivity to morality - a thinking entity that simply aimed to represent the world truly but which was not a practical agent, might need physics, but would have no need of moral beliefs unless she was actually a practical agent. (Korsgaard 2003)

However, Kantian constructivists think that the problems morality attempts to solve are those that must be confronted by *any* rational agent in the practical exercise of her agency. That's not to say that all rational agents actually find these problems to be significant - they might deny that morality provides solutions to the problems they care about. But if they were only to apply their powers of rationality more carefully, they would indeed come to care about practical problems to which the precepts of Kantian ethics provide a solution. Therefore, for the Kantian, all rational agents are within the persuasive scope of moral reasoning. Any rational agent, could, in principle, be persuaded to be moral.

If that were true, moral reasoning would have incredible power and significance. The arguments that philosophers give would have the potential, in principle, to rid the world of wickedness, if only the knaves would listen. The standard of moral correctness would not quite be objective, in the absolute sense; but it would still be universal, in the sense of being available to all rational agents.

Perhaps there are problems that all rational agents as such face. And perhaps solving these problems does indeed cause us to take on all the commitments and values that we normally think of as the "ethical" ones. But it is not obvious that this is the case. It does not strike me as odd if we have to say that there are some rational agents who can merely be condemned as brutes or opposed as forces of nature, but whom we cannot possibly hope will be genuinely moral agents. Whatever power or authority we assign to moral reasoning, it seems to ask too much to require that it be capable on its own of curing the world of the evils wrought by cruelty, callousness and indifference. Perhaps some rational agents are outside ethics altogether - capable of practical thought and deliberation, but

simply incapable of caring about the things that give morality its persuasive power. Moral reasoning could have quite enough authority to count as weighty and significant even if its potential audience was more limited than that of all actual and possible rational agents.

Convergence and Consensus

Michael Smith argues that our “preoccupation with moral argument” only makes sense if we attribute great persuasive powers to reasoned moral argument (Smith 1991 p1999). However, he seems happy to draw the circle of moral persuasion more tightly than Korsgaard - it need not be rational creatures as such that are capable of being persuaded; what is important is that all *humans* can be persuaded. Moral thought involves desire, not just pure reasoning, but the hope is that the desires of humans, so long as they are “cool, calm and collected” (p 406) can be corrected through rational thought and argument in order to yield a consensus. As he argues:

What seems to give moral arguments their point and poignancy is the idea that, since we are all in the same boat, a careful mustering and assessment of the reasons for and against our moral opinions is the best way to discover what the moral facts really are. If the participants are open-minded and thinking correctly, such an argument should result in a *convergence* in moral opinion - a convergence upon the truth...The term “objective” here simply signifies the possibility of a convergence in moral views of the kind just mentioned. (p199)

Of course, Smith is no constructivist or pragmatist. His notion of correctness is not the pragmatist’s conception of solutions to practical problems. Rather, as can be seen here, he thinks that convergence would vindicate the moral realist’s claim that there are mind-independent moral truths to which moral claims correspond. Further, he thinks that convergence would make morality objective. The inference from the fact of convergence to the existence of a mind-independent moral reality is certainly contestable. If we withhold that inference, it would be better to say that the moral perspective created by convergence is an *inter-subjective* rather than objective one. And perspectival inter-subjectivity is normally seen as an anti-realist, rather than realist, position in ethics (Sayre-McCord 1986).

Furthermore, Smith holds the neo-internalist view that, if convergence is not possible, then we have to accept that there is a “*fundamental relativity* in what reasons we have”. Here he is not using reasons-talk in Williams’ limited sense, but rather views reasons-claims as fundamental to all normative thought. So he means to say that the normative authority of morality would be dependent on having the right desires; if cool, calm and collected agents cannot be brought to care about the same things that a morally decent agents care about, then they cannot be criticised for that. Again, I reject this kind of inference, as explained in the previous section.

But my real interest in discussing Smith is the idea that we need to hold out hope for total moral convergence in order to make sense of our interest in moral argumentation. *That* hope can, after all, make sense even outside of a robustly realist framework. It is the key dividing line, among those inspired by the Classical Pragmatist tradition, between the “messianic” (Levi 1998) followers of Peirce, whose conception of truth is “the opinion which is fated to be ultimately agreed to by all who investigate” (Peirce 1992, p. 139), and those inspired by Dewey, who are more open to the possibility that no such convergence is fated. Does the practice of moral reasoning presuppose that persuasion can bring all human agents to a Peircean point of convergence?

Certainly, if moral reasoning had an extremely limited power to persuade, that would seem to undermine our interest in it. Street’s “Humean” constructivism agrees with me and with Korsgaard that moral correctness is a matter of solving practical problems. But whereas I designate those problems as the ones facing mutually sympathetic social agents, and Korsgaard stipulates that they are the problems facing rational agents, with the implication that all rational agents share a core set of problems to which ethics offers solutions, Street is less restrictive.

For her, the relevant problems are simply *any* problems created by an agent’s desires. There is a place for reasoning here - means-ends and coherence-focused reasoning can bring an agent’s desires into

greater harmony.³⁷ But Street does not assume that all rational agents will find the same things to be practical problems. And, unlike Smith, she doesn't assume that, cool, calm and collected humans will necessarily be able to converge in their desires. She thinks it unduly optimistic that we can find a common set of practical problems to serve as a shared basis for moral deliberation - desires are simply too diverse, and cannot be brought into line simply by moral reasoning. Convergence is not even on the cards.

Like Korsgaard and Smith, Street accepts the neo-internalist view that the authority of morality is contingent upon its ability to persuade agents through reasoning. So Street accepts an extreme form of normative relativism - what agents ought to do and can be criticised for is entirely contingent on their desires. The ethical world is massively fractured - for each set of desires there is a set of solutions which is normatively authoritative for the agents who have those desires, and not for others. We can criticise other agents for failing to solve their own problems. But if their problems are not the same as ours, we cannot criticise them for failing to adopt the same moral standards as us. A "perfectly coherent Caligula" stands beyond the boundaries of moral criticism. (Street 2009)

The world that Street describes would be a dreary one. If there were no common basis for practical deliberation, then moral reasoning would have extremely little power. Certainly, we could persuade people to be more coherent and instrumental in the pursuit of their own idiosyncratic goals. But I see little interest in this; why bother with moral reasoning if it could only persuade Caligula to be *more* coherent in his cruel perversions? Other things being equal, I would prefer a *less* coherent Caligula. If we deny neo-internalism, we can sensibly reject Street's normative relativism - there is no reason for *us* to endorse or refrain from condemning Caligula's activities, even if they are perfectly coherent. And we do not have to see the problems arising from just *any* desire as creating a standard of moral correctness. But still, if human desires were really massively diverse and

³⁷ Although whether that is always worth doing is debatable - see Hayward 2017.

irrevocably conflicting, if the world were filled with such Caligulas, then I do think that moral reasoning would largely be a waste of time. It would have little role to play in making the world a better place.

Fortunately, I do not think that the world is like this. I don't suppose that all agents, not even all human agents, are sympathetic and social.³⁸ But I think that many of them are. As such, I think that the community of agents who are capable of being motivated to take the problems I designate as moral, and hence with whom I can engage in persuasive argumentation, is in fact very large. In fact, I think that most of the world's population share enough with me that I can sensibly engage them in moral argumentation, with the outcome that they not only get better by their own standards, but also by mine. And they can hope for the same thing.

But this hope of mutual engagement, and perhaps of gradual convergence, is not the same as movement towards a single, Peircian limit point. It might be that our shared sympathetic sentiments give us the possibility of finding extensive common ground, but that the diversity of our other motivations makes absolute consensus unachievable in ethics. It might be that whilst we can agree on shared standards for living together in certain public arenas, there will be many other areas of ethical thought and practice where different communities and sub-communities settle on quite diverse norms and principles. And perhaps the shared problems that ground mutual engagement admit of a variety of equally good solutions, so that there is no one unique convergence point at which even the mutually sympathetic are aiming. That's to say, perhaps a variety of different sets of ethical norms, principles and practices could solve the problems sympathetic agents find in living together, and are thus equally correct. Again, that means there is no one point of convergence. And it strikes me as quite possible that the future of human social life will continue to produce new

³⁸ Although Street is quite explicit that she is considering hypothetical agents whom she takes to be quite different from normal humans - real Caligulas are not ideally coherent, and might have some internal impetus in favour of moral reformation.

problems that are quite unlike those we have seen in the past, requiring us to engage in endless moral innovation to create entirely novel solutions. In that case, there is no Peircian limit point because there can be no end to moral inquiry.

As Smith argues, I think the possibility of convergence is something to be demonstrated in the course of history, not established in advance through *a priori* argument. All the scenarios I just mentioned seem quite plausible to me, but I do not know for sure that any of them is in fact accurate. Yet unlike Smith, I do not see these possibilities as a threat to the authority of moral reasoning. Even if there is no unique, fixed, homogeneous consensus point to which moral reasoning will take us all, reasoning still has a point.

For I think the scenarios above are entirely appealing - they do not seem to be to describe a worse, more dreary and disappointing moral world than the world of absolute consensus. Moral variety and moral change, within limits, seem to me quite appealing. Indeed, perhaps a world of moral diversity is even preferable to that of absolute consensus. As Mill and his followers have pointed out, it would certainly be offer a greater wealth of resources in dealing with new problems. A society in which many different sub-communities pursued “experiments of living” with different rules and norms would be more adaptable in the face of moral change - perhaps the society at large might find the practices of a minority worth adopting in the face of new challenges. (Mill 1989; Kitcher 2011) As such, I think that moral reasoning would be quite worth engaging in even if it could not bring about moral consensus; bringing about these other scenarios is just as worthy a goal.

Tribalism and Humanism

Even if the world is full of sympathetic and social creatures, there is a possibility that, if true, would seem to threaten our understanding of ethics. For I have said nothing so far about the limits of sympathy. If human beings were only able to extend their sympathies to those in their family, kin-

group, tribe or nation, then the power of moral reasoning would be worryingly limited. We might work out with our fellow tribesmen how best to solve the problems that we find in living together. But if we could not reach out across such boundaries then that would be a tragic situation - there would be no possibility of creating a normative community that stretched across families, tribes or nations. Beyond those boundaries, there would be no possibilities beyond indifference, war or uneasy truce.

If purely tribal ethics were the best that we could hope for, then that would not be a reason to deny authority to ethics altogether. Solving the collective problems of a tribe or kin-group is an important problem, and so a practice that attempted to do this would be worth respecting. We can see this by considering a world in which a purely tribal morality were the only possibility.

Darwin considered what morality would have emerged had “men [been] reared under precisely the same conditions as hive-bees” (Darwin 1871, p70). Since we are interested in deep-seated psychological and motivational capacities whose lineages no doubt stretch far back in evolutionary history, let us instead imagine a race of highly intelligent creatures descended from actual bees, rather than, like us, from great apes. They would have an incredible capacity for altruism towards other members of the same swarm. But they might be entirely incapable of concern for bees from other swarms. Once they evolved to a level of sophistication in which they could identify and collectively solve practical problems, they would focus most on collective survival and flourishing, and create norms that required limitless personal altruism. But their codes would be silent about the treatment of foreign bees, or even counsel merciless aggression. They would recognise no *shared* problems with other swarms which might admit of collective solution.

I am not entirely sure how we should evaluate the morality of these bee-men. I think we could acknowledge the value of their social norms, and even afford them a certain admiration for their

feats of mutual altruism and care. That is to say, we would not entirely condemn them, and might agree that their norms had just the same authority over them as ours do over us. But perhaps we might nevertheless despise them for their inability to reach beyond the bounds of their swarms and form an international community of bees, to collaborate in the creation of norms for all bee-men regardless of their origins.

Fortunately, however, men are not bees. Perhaps we start with “no such passion in human minds, as the love of mankind, merely as such” (Hume 2000 Book 2, Part 3, Section 3) - that is to say, with sympathies and pro-social sentiments that are severely limited in the manner described above. And so long as individuals are not in frequent contact with strangers, then a purely tribal morality needn't be condemned as entirely defective. But fortunately, we are able to extend our sympathies very far from this beginning. There is no better argument for this claim than the fact that it has actually happened. The process of identifying collective problems in diverse societies and across the world's nations, of recognising the interests of others as worth pursuing, and of using moral reasoning to solve these problems has progressed, albeit fitfully. At least some of us care enough about those who are entirely unrelated to us to see ourselves as having collective problems, and to see these as having normative authority over us. Many philosophers have viewed the history of moral development as one of an “expanding circle” (eg Singer 1981). I disagree with the view that all moral improvements can be captured in terms of this single notion.³⁹ But I do think it is true that the circle of concern has indeed expanded at various times in human history. This is because our sympathies are not tribal, but, for want of a better word, humanistic. Even if we do not *start* with a love of mankind as such, we are capable of developing one.

How much farther might the circle expand? How much farther should we *want* it to expand? Could morality aim to solve the problems that *all* animals find in living together, or is it limited to humans,

³⁹ A sympathetic critique of this idea is found in Kitcher 2011 pp214-217. Kitcher sees the expanding circle as one form of moral “progress”, but not the only one.

or perhaps humans and a handful of higher animals with similar psychological capacities? What exactly would it mean to adopt a pragmatic conception of correctness that embraced the interests of animals that could not, themselves, grasp norms and principles, or offer mutual sympathies to us? These are important questions, and indeed beset the elaboration of any constructivist or pragmatist view. But I leave them for another day. So far, our circle has expanded a long way. Perhaps it would be better if it could expand even further. But we should not despair or reject morality if it cannot, just as we should not entirely condemn the purely tribal morality of the bees. Even if morality has limits, we should recognise its authority and see the value of moral reasoning.

Conclusion

Mine is a view that dispenses with much that has often been seen as necessary for us to accept the authority of morality. An objective perspective, rational sanction, universally motivating moral arguments, and the possibility of moral consensus are not, I have argued, necessary for morality to matter. Some might see my views as pessimistic. Would it not be wonderful if we were able to refute the knaves and use philosophy to purge the world of immorality, to encompass all rational agents and creatures capable of objectivity in one moral community, to pursue convergent lines of moral inquiry to a single point of final, unshakeable consensus? Perhaps. But these are grand ambitions, and it is far from clear that the world will co-operate in their realisation. To me, they have the ring of Holy Grails. As Blackburn says, we should not hold morality hostage to our philosophical ambitions. And, as I have argued, we do not need to. This is a deeper form of optimism. Even in a humdrum world devoid of these sparkling trophies, morality is worth caring about.

Bibliography

Blackburn, Simon (2010) "Must We Weep for Sentimentalism?" in *Practical Tortoise Raising* (Oxford, Oxford University Press)

Bloomfield, Paul (2009). "Archimedeanism and Why Metaethics Matters" in *Oxford Studies in Metaethics Vol 4*, ed. Russ Shafer-Landau (Oxford: Clarendon Press)

- Buckwalter Wesley & John Turri (2015) "Inability and Obligation in Moral Judgment" *PLOS ONE* 10(8): e0136589. <https://doi.org/10.1371/journal.pone.0136589>
- Burgess, John (1983) "Why I am Not a Nominalist" *Notre Dame Journal of Formal Logic*, 24: 93–105.
- Carnap, Rudolf (1950) "Empiricism, semantics, and ontology" in *Revue Internationale de Philosophie* 4 (11):20--40 (1950)
- Chituc, Vladimir, Paul Henne, Walter Sinnott-Armstrong, & Felipe De Brigard. (2016) "Blame, not ability, impacts moral 'ought' judgments for impossible actions: Toward an empirical refutation of 'ought' implies 'can.'" *Cognition*, 150(C), 20–25
- Cuneo, Terence & Russ Shafer-Landau (2014) The Moral Fixed Points: New Directions for Moral Nonnaturalism in *Philosophical Studies* 171 (3):399-443
- Darwin, Charles (1871) *The Descent of Man* (New York: Appleton)
- Dasgupta, Shamik (MS) "Non-Naturalism and Normative Authority"
- Dworkin, Ronald (1996) "Objectivity and truth: You'd better believe it" in *Philosophy and Public Affairs* 25 (2):87–139 (1996)
- Enoch, David (2006) "Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action" *The Philosophical Review* Vol. 115, No. 2 (Apr., 2006), pp. 169-198
- Greene, Joshua (2008) "The Secret Joke of Kant's Soul" in W. Sinnott-Armstrong ed., *Moral Psychology*, Vol. 3. (Cambridge MA: MIT Press)
- Griffin, James (2015) *What Can Philosophy Contribute to Ethics?* (Oxford: Oxford University Press)
- Joyce, Richard (2006) *The Evolution of Morality* (MIT Press, Cambridge MA)
- Joyce, Richard (2016) "Moral Fictionalism" in *Essays in Moral Scepticism* (Oxford, Oxford University Press)
- Hayward, Max Khan (2017) "Practical Reason, Sympathy and Reactive Attitudes" *Nous*. doi: 10.1111/nous.12203
- Hayward, Max Khan (MS a) "Nonnaturalist moral realism and the limits of rational reflection"
- Hayward, Max Khan (MS b) "Immoral Realism"
- Hume, David (1994) *An Enquiry Concerning the Principles of Morals* L.A Selby-Bigge ed, revised P.H Nidditch (Oxford: Clarendon Press)
- Hume, David (2000) *A Treatise of Human Nature* (Oxford: Oxford University Press)
- Hussain, Nadeem & Nishi Shah (2006) "Misunderstanding metaethics: Korsgaard's rejection of realism" in Russ Shafer-Landau ed., *Oxford Studies in Metaethics* 1:265-294 (Oxford: Clarendon Press)

- James, Aaron (2012) "Constructing Protagorean Objectivity" in J. Lenman & Y. Shemmer eds., *Constructivism in Practical Philosophy* (Oxford University Press)
- James, William (1956) "The Moral Philosopher and the Moral Life" in *The Will To Believe* (New York)
- Kahane, Guy (2011) "Evolutionary Debunking Arguments" *Noûs* 45 (1):103-125 (2011)
- Kitcher, Philip (2011) *The Ethical Project* (Oxford: Oxford University Press)
- Korsgaard, Christine (2003) "Realism and Constructivism in Twentieth-Century Moral Philosophy" *Journal of Philosophical Research* 28 (Supplement):99-122.
- Kramer, Matthew (2009) *Moral Realism as a Moral Doctrine* (Hoboken NJ, Wiley-Blackwell)
- Lenman, James (2012) "Expressivism and constructivism" in J. Lenman and Y. Shemmer eds., *Constructivism in practical philosophy* (Oxford : Oxford University Press)
- Levi, Isaac (1998) "Pragmatism and Change of View" *Canadian Journal of Philosophy Supplementary Volume* 24 (1998): 177–201
- Mackie, J.L. (1977) *Ethics: Inventing Right and Wrong* (London: Pelican Books)
- Mil, J.S. (1979) "Utilitarianism" (Indianapolis, Hackett)
- Mill, J.S. (1989) "On Liberty" in *On Liberty and Other Writings* ed Stefan Collini (Cambridge: Cambridge University Press)
- Nagel, Thomas (1986) *The View From Nowhere* (Oxford University Press)
- Peirce, C.S., (1992) *The Essential Peirce Vol. 1* (Bloomington: Indiana University Press)
- Sayre-McCord, Geoff (1986) "The Many Moral Realisms," *Southern Journal of Philosophy* 24 (S1):1-22 (1986)
- Silk, Alex (2015) "Nietzschean Constructivism: Ethics and Metaethics for All and None" *Inquiry : An Interdisciplinary Journal of Philosophy* 58 (3):244-280.
- Singer, Peter (1981) *The Expanding Circle* (New York: Farrar, Straus, Giroux, 1981).
- Singer, Peter (2005) "Ethics and Intuitions" *Journal of Ethics* 9 (3-4):331-352 (2005)
- Smith, Michael (1991) "Realism" in Peter Singer, ed., *A Companion to Ethics* (Oxford: Basil Blackwell)
- Street, Sharon (2006) "A Darwinian Dilemma for Realist Theories of Value" *Philosophical Studies* (2006) 127:109166

Street, Sharon (2008) “Constructivism About Reasons” in Russ Shafer-Landau ed. *Oxford Studies in Metaethics* 3:207-45 (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)

Väyrynen, Pekka (MS) Normative Commitments in Metanormative Theory

Williams, Bernard (1978) *Descartes: The Project of Pure Inquiry* (Harmondsworth: Penguin)

Williams, Bernard (1981) “Internal and External Reasons” in *Moral Luck* (Cambridge: Cambridge University Press)

Williams, Bernard (2006) “Philosophy as a Humanistic Discipline in *Philosophy as a Humanistic Discipline* (Princeton University Press)

Conclusion

I have argued that, from the moral point of view, we should abandon the demand for objectivity, universality and mind-independent truth in ethics. To see ethics as created by us, to solve the kinds of problems we face in living together and caring for one another, vindicates the authority and significance of ethics. This is the humanistic claim, a claim about what matters.

But that doesn't mean that it is easy to live without objectivity. It means abandoning a search that has compelled and motivated generations of ethicists, and facing up to hard truths. James famously distinguished between the tough- and the tender- minded tendencies in philosophy; anti-realism and sentimentalism in ethics clearly tend to the side of the toughs. James saw pragmatism as a compromise between the two traditions, but it would be better to see his project as a starting from a tough-minded position and finding a way to acknowledge and accommodate the intellectual and emotional needs of the tender-minded within that basic framework. (James 1907) In that spirit, it is worth taking a moment to reflect on what is lost in accepting my picture of the ethical world, and suggesting what it would mean for moral philosophy to confront that loss.

First, it means giving up on the moral conversation-stopper of absolute refutation, the Holy Grail identified by Blackburn that I addressed in papers 3 and 4. There may be rational agents, capable of objectivity, who simply cannot be compelled by any argument to respect the authority of morality. And indeed, it is not only those who are emotionally incapable of moral motivation that lie beyond the scope of philosophical persuasion. Real, human agents who have not yet extended their sympathies in the manner described in paper 4 are (usually) *capable* of doing so, and of adopting the moral perspective. But it takes more than reason thus to draw them out. That is disappointing for a

philosopher - many of us see it as within our purview to foster and sharpen our students' capacities for rationality and objectivity, but our training leaves us ill-equipped to effect emotional development and nurture sympathy.

To acknowledge the limitations of philosophy in bringing about moral reform would mean to look to beyond our discipline; to the social sciences, and to the other arts and humanities. Just because rational argumentation and philosophical acuity don't, unaided, have the power to reform the immoral, or even just the morally stunted, it doesn't follow that nothing has that power. A moral philosophy better integrated with empirical research into the actual factors that promote a sympathetic enlargement of the humane sentiments, and into the stories, narratives and cultural productions that seem so well-suited to effecting emotional development, would have greatly expanded powers to contribute to the practical business of solving human problems.

But the renunciation of conversation-stoppers doesn't just affect our relationship with others. It means giving up on another kind of finality - the finality of closing a difficult topic. The urge to pursue philosophy is not always a pleasant one, but can rather spring from an irritation of doubt. It would be nice to get cleared up on questions in the foundations of value, to figure out what is ultimately good, right and valuable, in order to get on with other, more immediately practical work in applied ethics, politics and public policy. Oftentimes it is a desire to make a difference in the latter areas that drew us to ethics in the first place - in that case, nagging worries about foundational questions seem more frustrating than interesting. Many philosophers like to think that philosophy is something that can be finished and surmounted, in one of two ways - either because philosophical questions admit of final resolution, or because philosophy is somehow mistaken and philosophical worries require therapy rather than constructive thought.

If I am right, then there is no prospect of “finishing” ethics in either of these ways. The foundational questions will continue to plague us, because each generation will have to ask whether the moral conceptions they have inherited really do a good job in solving the practical problems they confront - problems which are themselves evolving and changing over time. These are real questions, but not ones that admit of final resolution. And so moral philosophers will never be able to finish the irksome work of questioning the foundations of their value-system.

But the view I propose should make the contrast between “practical” questions of applied ethics and policy and “theoretical” questions about the foundations of value seem less acute, and hence the lure of the former less pronounced. Moral philosophy is not just a matter of discovering facts that are there already in order to apply them to practical situations, but involves actually inventing principles and norms with a view to their function. Asking foundational questions really can make a difference.

It may seem obscure how this can have a practical effect on a society like ours - whereas academics are occasionally consulted on questions of, say, clinical ethics or social policy, it is not clear how philosophers can change the norms governing an entire culture or society. Yet this should not seem like an insurmountable challenge. The business of critiquing and reforming our collective moral standards is ongoing in our society; what is needed is for philosophers to join in the conversation. Thus, like all pragmatists, I think that a proper appreciation of the nature and purpose of philosophical thought requires that philosophers take on a greater role as public intellectuals.

My view, if accepted, also has challenging implications for our sense of moral responsibility. Often obeying the strictures of morality requires us to do things that are, on the face of it, quite unpleasant - to impose costs or punishments on those who are deemed to deserve it, or to enjoin others to endure pain and sacrifice, or to determine that the limitations of an individual’s right to aid and

protection. This is why many philosophers are unwilling to intervene in the practical world until they have assuaged their moral doubts - the costs of being wrong are high.

But there is still something reassuring about the thought that, when we impose the unpleasant parts of moral life on ourselves and others, we are simply obeying prescriptions that have their source quite outside of ourselves; likewise, many people find it comforting to think that they are just “obeying orders” when they have to perform acts they find unsavoury. If moral norms really are something that we create, then we cannot palm off responsibility for them in this manner. We are ourselves ultimately responsible for morality, for its unpleasant as much for its appealing parts, and that is an emotional burden to bear. But I think that this burden is something we should embrace, since it is simply the counterpart of accepting how important the work of inventing values is; if it weighs philosophers down with a greater seriousness as they construct their arguments and thought experiments, then that is only a good thing.

Finally, accepting my kind of anti-realism robs us of a salve for a certain kind of existential worry. It is easy in moments of despair or misanthropy to feel disappointed that the merely human natural world is *all there* is; how could mere human feelings be a source of meaning and value? When philosophers like Parfit worry that nothing matters, I suspect that they are giving philosophical voice to existential worries of this kind. And I understand how this can make the search for objectivity compelling. The thought that there is a glittering realm of values just *out there*, waiting to be discovered, purified of human particularity, necessary and universal, can reassure us in such dark moods. Ethics could hardly help us find meaning at times when we feel incapable of sympathy and human sentiment if ethics is *itself* a product of sympathy and sentiment.

Again, my only counsel is that we have to recognise the limitations of philosophy here. Just as there is no conclusive proof or irrefutable argument that can compel the amoralist, so there is no product

of mere philosophy that can insulate us from despair. The man who worries that nothing matters, who cannot bring himself to see human problems and purposes as giving meaning to life, does not need an argument or demonstration. Perhaps art and literature might help him escape that bleak perspective; perhaps he needs love, or medicine, or meditation. Perhaps nothing can help him. Confronted with the question of whether life is worth living, James offered no better response than “Believe that life is worth living, and your belief will help create the fact” (James 1895 p25) - an answer that could hardly have helped the man who had decided that life is *not* worth living. If values are likewise created, we may have no salve to offer to someone who cannot see how anything can matter in the merely human natural world.

I want to finish with a few words about what I have left undone in these papers, and what I aim to do next. First, whilst many of my arguments take aim at robust forms of non-naturalist realism, my engagement with other forms of moral realism is relatively perfunctory. There is a reason for this - whereas non-naturalism is a relatively clear and homogeneous doctrine, and thus a stable target for criticism, the multiplicity, variety and complexity of the other views that fly under the flag of “realism” precludes any simply exposition. I have gestured at the claim that certain views, like Cornell realism and particular forms of quasi-realism, might be susceptible to my critical arguments; other approaches, like Putnam’s later moral philosophy, may belong on my side of the dividing line, despite their assumption of the realist label. Whereas Parfit hoped that all metaethical theories were “climbing the same mountain”, differing only in the routes they took to a single summit, I suspect that there might be two mountains. That’s to say that the key insights of naturalism, constructivism, pragmatism and expressivism may all be compatible, and compatible in a package that stands opposed to rationalism, objectivism and realism. But this is an argument that I have not made yet.

Secondly, my own view has only been specified in its outlines. Two questions strike me as standing in particular need of clarification. First, I have said little about what it takes for something to count as a practical problem, and how, given any problem or set of problems, we are to go about identifying a solution. I suspect that it will be impossible to give a neat general answer that goes beyond the bare and unhelpful claim that there are human problems when there are humans who find themselves to have a problem. Rather, our identification of problems will be “theory-laden” in the same manner that scientific observations are theory-laden. We can see science as being in the service of certain general goals whilst recognising that the precise articulation and specification of those goals can only be performed in the language of the sciences of the day, and hence will stand open to revision in the face of scientific change. Likewise, whilst we can accept the general claim that ethics is valuable because it solves human problems, the tools for identifying problems and solutions will be part of our practical ethical outlook, and hence will themselves need to be scrutinised over time.

More detail is also needed in spelling out the precise moral psychology that I have in mind. I say that the moral problems are those that mutually sympathetic affective agents find themselves to have. But what affects or emotions do these agents need to have in order to count as ethical agents; and what conception or conceptions of sympathy do I have in mind? I have taken my cues largely from the classical sentimentalists, Hume and Smith, and I think that their descriptive claims about human psychology remain influential in contemporary empirical psychological research. But still, I aim to say more about the psychological mechanisms that I have in mind, and how these relate to our current understandings of the moral emotions.

Third, the papers presented here have focused primarily on *moral* values. But there are other forms of value and normativity in philosophy, and some of my arguments might seem naturally to extend beyond the boundaries of ethics. If ethics and practical rationality gains authority only because it helps us to solve our collective problems, how could the normativity of epistemology or theoretical

rationality have any other source? Indeed, I do intend to extend these arguments in precisely the manner so indicated - sentimentalism is not just a theory about ethics, but about normativity in general.

Finally, we might wonder about whether any of this really has any practical significance. If sentimentalism identifies the heart of ethical normativity in collective human problems, does it offer us any particular resources in solving those problems? Is there a distinctive approach to practical normative philosophy that stems from my metaethical views? I think that there is. But the articulation of that approach must wait for a future project.

Bibliography

James, William (1895) "Is Life Worth Living?" *International Journal of Ethics*, Vol. 6, No. 1 (Oct., 1895), pp. 1-25

James, William (1907) "The Present Dilemma in Philosophy" in *Pragmatism: A new name for some old ways of thinking* (New York: Longman Green and Co)