DISENTANGLING THE WILL[1]

RICHARD HOLTON, MIT

Nietzsche held that the idea of the will—and he meant specifically free will—has its unity only as a word.[2] That is perhaps to go too far, but in spirit at least I am inclined to agree. There are at least three ideas bundled up in the idea of free will and I doubt that there is any one thing that fits them all.

First, there is something that has its place in the theory of mind and action. Freedom of the will is the property—more exactly the capacity—possessed by agents who are able to act freely, a capacity that is manifested whenever they do act freely. Since we are talking about a mental capacity here, we should expect to find out about it using our normal tools for finding out about mental phenomena: using the empirical resources of psychology and neuroscience, and the conceptual resources of philosophy of mind. But a good starting point, one that is particularly useful for getting a grip on what it is that we are talking about, is through the phenomenology of agency. We have some knowledge of free will in this sense because we have a direct experience of it.

Second, there is a notion that is distinctively moral. Free actions are actions for which agents are morally responsible; that is, freedom is sufficient for moral responsibility. Perhaps, in addition, freedom is necessary: perhaps every act for which an agent is morally responsible is an action that they perform freely. Investigation of this issue is a task for ethics and for moral psychology.

Third, there is a modal idea that has its natural place in metaphysics. Freedom of the will is that property possessed by agents who, when they act, could have acted otherwise. Understanding quite what this 'could' amounts to has proved a difficult task, especially if the world is deterministic at the macroscopic level. But that is a task for metaphysics and semantics, and perhaps for physics. It is implausible that this is something that we are aware of directly.

To say that these are separate ideas is not to deny that they interact in very many ways. But it is a substantial claim that there is a single thing that does duty in all three. The familiarity of the claim should not blind us to the fact that it is not forced upon us. It could well be, for instance, that the actions that we experience as free do not exactly overlap with actions that our best moral theory will tell us are those for which we are responsible. I shall argue that this is indeed the case, and moreover that something stronger is true: that even when we generalize the mental conception of freedom to

---

[1] This paper was presented at the Conference on Agency and Responsibility, University of Indiana 2007. My thanks to the audience there, to Tim O'Connor who was my commentator, and to Eddy Nahmias and Rae Langton for comments on the written version. Many of the ideas discussed here are developed further in a recent book (Holton, 2009).

[2] (Nietzsche 1886) §19.

extend beyond those actions that we are immediately aware of as free, we still do not get something that is quite right for our theory of moral responsibility. The mental notion of freedom is distinct from the moral one; and the modal one, I suspect, is something different again.

This approach becomes more plausible when we reflect on the history of our understanding of freedom. The claim that there is one thing that can play all three roles —mental, moral and modal—is a claim took many hundreds of years to evolve. There is considerable scholarly controversy over who first made it. Richard Sorabji constructs a persuasive case that it was not present in classical thinkers, and first really comes together in the work of Augustine.[3] Of course, that was still a very long time ago, and the influence of thinkers like Augustine, transmitted as it has been by the Christian Church, has been tremendous. It might be held that nowadays our *concept* of free will is essentially the concept of some one property that can play all three roles. On this view, to deny that there is anything that can do so is exactly to deny that there is free will.

I'm not sure what to say about our *concepts*. Or, more accurately, I'm sceptical that there is anything very helpful to be said. There has been a flurry of experimental work recently concerned with asking subjects their views about free will: asking whether, for instance, there would be moral responsibility in a world that was deterministic.[4] Such work is interesting and important, and is certainly a great advance on simply asserting what people's ordinary intuitions are, without troubling to find out. But how much should we conclude from these experiments about the nature of our concepts? Suppose that a majority of subjects hold that a deterministic world would be one in which moral responsibility were absent. Should we conclude that our concept of moral responsibility requires the falsity of determinism?

I think not. Around 300 years ago, Calvinism was a major force within Christianity. In large parts of Germany, the Netherlands, Scotland and New England it was the dominant doctrine. If you had asked congregations in those places whether moral responsibility was compatible with determinism, I think there is little doubt that they would have said that it was: predestination was a central tenet of their beliefs. But should we say then that these people had—or contemporary Calvinists have—a different concept of moral responsibility to that of the subjects who answered differently in our experimental surveys?[5] Or should we rather say that Calvinists and

---

[3] (Sorabji, 2000). Sorabji distinguishes even more elements to the notion than I do.

[4] See, for instance, (Nahmias *et al.*, 2005) (Nichols and Knobe, 2007)

[5] It might also be held that Calvinism is simply incoherent, exactly because it is incompatible with our moral beliefs. This contention could then be supported by the observation that it has lost ground, within the Protestant churches, to Arminianism. Indeed, even in those churches that are nominally Calvinist, the significance of the doctrine of predestination has surely declined—I was brought up a Congregationalist, but didn't realize through thirteen years of Sunday School that this was a doctrine to which the church was committed. Nevertheless, Calvinism is far from a spent force, and there may be other reasons why it currently has less of a public profile. My father-in-law once told me that as a trainee missionary he was advised to be Arminian in his preaching, but Calvinist in his prayer.

Arminians have different theoretical views about the same concept? Certainly the great architects of the view—Luther and Calvin, and many since them—thought that they were coming to a better understanding of moral responsibility and the role of the will, not replacing these notions with others.

If Twentieth Century philosophy has taught us anything, it is that such debates are sterile. We have no clear criteria for what counts as an essential part of a concept, and what counts as a detachable theory about that concept. Certainly this is true: our intuitions about freedom stem from many sources, amongst which are theoretical views about what roles freedom must play. If, as I want to suggest, our notion of free will has run together disparate things that would be better kept apart, then there will be even less chance of distinguishing any essential core of our concept from the ideas we have about it.

Let me give a second example. In a fascinating study, Eddy Nahmias and colleagues found that people think that neurological determinism (everything we do is determined by the prior arrangement of our neurons) poses more of a threat to free will and to moral responsibility than does psychological determinism (everything we do is determined by our prior beliefs, desires and intentions).[7] Again we might take this at face value to show something about our ideas of free will and moral responsibility. But there is something curious about the finding. For, on broadly physicalist assumptions—that is, assuming that everything supervenes on the physical—the thesis of neurological determinism is *weaker* than the thesis of psychological determinism. Neurological determinism could be true whilst psychological determinism is false: it could well be that many of the neurological mechanisms influencing our behaviour work without giving rise to psychological states like beliefs, desires and intentions. But, given the truth of physicalism, it is very unlikely that psychological determinism could be true whilst neurological determinism were not; it is very implausible that psychological states are realized in anything other than neurological states.

So how can we explain Nahmias's findings? The obvious suggestion is that, at least implicitly, the subjects are rejecting physicalism. If they were dualists—if they thought, like Descartes, that mind and matter were quite distinct, so that each could vary independently of the other—then the results would make perfect sense. Neurological determinism is incompatible with all but the most recondite versions of dualism, since neurons are clearly physical, and so, given dualism, cannot determine the mental. In contrast, psychological determinism is quite compatible with dualism. So plausibly it is a set of (perhaps implicit) dualist assumptions that are leading the subjects to react as they do. If that is how to explain the findings though, then it does not simply show that

---

[6] The classic work is (Quine 1951); but work by Saul Kripke and Hilary Putnam, work denying that natural kind terms have much descriptive content, has been equally influential on this point; see (Kripke 1972) and (Putnam, 1975).

[7] (Nahmias et al, 2007)

the subjects are dualist. More significantly, it shows that they think that free will stands or falls with the doctrine: free will is possible with dualism, impossible without it.

That is not a terribly surprising finding, especially for subjects, like the Georgia undergraduates who participated in this survey, who have been strongly influenced by Christianity. On a standard understanding, Christianity is clearly a dualistic system, and I would wager that a good number of those involved would think that there is no morality without Christianity. So, on that view, if dualism is false, Christianity must be false, and there will be no morality. But even if it is not terribly surprising that most of these subjects think that dualism is necessary for free will, how much should we conclude from it for our concepts of free will? Dualism has had something of a renaissance in recent years, but it is still very much a minority view amongst philosophers, and even more so amongst most scientists who work on the mind. Free will would be a much less plausible thing than it is generally taken to be if it required the truth of dualism. Again I would be reluctant to draw any firm lines, but I suggest that we should say that these findings show at least as much about the beliefs that some subjects have about free will, as about the concepts themselves.

Indeed, when we look more carefully at the studies of what ordinary subjects believe about free will we find that they show a very mixed picture. Get them to think about the issue in the abstract, and most people do think that moral responsibility would be absent in a physically deterministic world.[8] But that finding is reversed once the subjects think about a concrete case: get them to focus on some particular nasty individual, and they will think he is responsible even in a deterministic world.[9] Equally, get them to think, not of how they would react to some other possible world in which determinism is true, but of how they would react if they discovered that determinism is true of our world, and again they now think that responsibility and determinism are compatible.[10]

I have argued that our ideas about free will come from diverse sources, and that we should not always take intuitions about free will as indicating the nature of the properties involved, since they may simply reflect false theory about those properties. But if this is right, how are we to make any progress? I don't advocate a single alternative method, since it seems to me that we should be open to considerations from any number of different areas. However, in this paper I do want to highlight two approaches, corresponding to two of the three sources of our concept of freedom that I mentioned at the outset. When it comes to the philosophy of mind, I think that we should be more carefully attentive to the phenomenology of freedom that we have customarily been. When we do, we will find that there are at least two different experiences to which we should pay heed: an experience of choice, and an experience of

---

[8] See, for instance, (Nichols & Knobe, 2007)

[9] (Nahmias *et al.*, 2005)

[10] (Nichols and Roskies, forthcoming).

agency (or perhaps, more accurately, of loss of agency).[11] I will argue that both can be seen as revelatory of real phenomena, but that neither is the central notion for moral responsibility. When it comes to ethics, my suggestion is that we should pay heed, not to our explicit moral beliefs, but to our actual moral practice. In particular, I want to investigate the idea of how we make attributions of moral responsibility to those who are mistaken about their own motivation.

About the third idea, the idea that freedom of the will is the ability to do otherwise, I shall not have a great deal to say; the enormous philosophical literature on the idea has shown just how hard it is even to work out what it means, let alone to give an account of it that is compatible with what physics seems to be telling us. However, my discussion of the other two ideas will indicate some ways in which it has come to feature in our thinking, and provide some reasons why we should not be over concerned about it.

Experiences of Freedom

Much of the force of the idea that we have free will comes from our experience. For some years now I have introduced undergraduates to the topic; and I find that the quickest, most effective way to generate the conviction that they have free will is to get them to focus on the phenomenology. Tell them to make an arbitrary choice, and then get them to act on it—to raise their left hand or their right, for instance—and they are, by and large, left with an unshakable conviction that their choice was a free one.

What is happening here? They have in the first instance an experience of freely choosing and acting. Quick on its heels comes a judgement, or a clutch of judgements: that they could have made either choice; or, more theoretically committed, that they could have made either choice compatibly with how they were prior to the choice; or more committed still, that they could have made either choice no matter how the whole world stood prior to the choice, and hence that they are, in that respect, unmoved movers.

Judgments like these last surely go well beyond the contents of the experience. How could one have experience that one's action was itself uncaused? Wouldn't that require that one also had experience of the rest of the world to show that it was not doing any causing? Nevertheless the experience of freedom is an experience of something. At its heart, I suggest, are two aspects. First, we have an experience that provides the basis for

[11] I think that at least one more experience is centrally important: the experience of making and maintaining resolutions. Like choosing this is something that we actively do, something that requires effort. I think that this explains why it is that inducing beliefs in determinism tends to undermine subjects' moral motivation; see (Vohs and Schooler, 2008) (Baumeister et al., forthcoming). Determinism is easily conflated with fatalism, with the doctrine that nothing one can do will make any difference to the outcome; as a result, belief in determinism can undermine moral self-efficacy. Subjects come to think that there is no point in trying to persist in any resolution to behave well, since their effort will have no impact on the outcome. For discussion see (Holton, 2009), Ch. 8.

a belief in the efficacy of choice, by which I mean that, once the question of what to do has arised, choice is both necessary and sufficient for action (choose to raise you right hand, and you'll raise it; likewise with your left; fail to make either choice and you won't raise either). Second, we have an experience of different choices being compatible with our prior beliefs, desires and intentions. Believing, desiring and intending as one does, one could either choose to raise one's left hand or one's right hand. In this sense we do have an experience that provides the basis for a belief that our actions are not determined: they are not determined by our beliefs, desires and intentions. But this local indeterminism falls far short of the global indeterminism that libertarians embrace. It is quite compatible with the thought that our actions are not determined by our beliefs, desires and intentions that they are nonetheless determined.[12]

When I say that the experiences *provide the basis* for these beliefs, I'm afraid I mean to leave the matter there; I am not going to pursue the difficult question, hard enough even in the case of ordinary perceptual experience, of the relation between the experience and the belief. We can think of our experiences of choice as broadly parallel to perceptual experiences, without, I hope, thereby committing ourselves too far. Like ordinary perceptual experiences, our experiences of freedom have special, infallible, authority. On the basis of experience we believe that striking the match is necessary and sufficient for its lighting; and, on the basis of our experience we think that taking the match out of the box is compatible both with its subsequently being lit, and with it never being lit. In such cases we might be wrong in the specific case—this match may not light even if it is struck, since it is damp; or (though this is far less probable) we may be wrong in general—we may be totally wrong about how matches work. Similarly, on the basis of our experience we think that choice is frequently necessary and sufficient for action, and that different choices are compatible with the same beliefs, desires and intentions. However, again it isn't ruled out by the nature of the experience that we be wrong about this. Again our error could be limited to the specific case—we think that we choose to raise our left hand, but really we are responding to post-hypnotic suggestion—or conceivably it could be more general—perhaps, as certain psychologists argue, choice is epiphenomenal.[13] I am, of course, committed to thinking that they are wrong about this, and that choice is choice is causally effective. But the reasons for thinking this cannot come just from the experience. The experience must be corroborated by a general account of how the mind works.

On this approach is an interesting question as to why we have choice. It can look like a liability, opening as it does the possibilities of *akrasia*—action against one's best

---

[12] The issues can get complicated here in spelling out how this would happen in a deterministic setting. For a start, it could be that beliefs, desires and intentions *together with* other independent states (e.g. choices) determine what agents do. But further, given that mental states are multiply realized (i.e. the same mental state can be realized in different neurophysiological states), it is quite compatible with physical determinism that two agents could be in the same psychological state, and yet would behave differently, since that state was realized differently in each of them.

[13] See especially (Wegner, 2002)

judgment—and inaction. Wouldn't we do better as creatures whose actions are linked directly to what we judge best, circumventing any need for choice? I have pursued this question elsewhere, and shan't address it in any detail here.[14] Briefly, my answer is that, as cognitively limited creatures, we are frequently unable to make a judgement as to what is best. We need to be able to choose to act even in the absence of such judgements. Such choices need not be random; they might instead be influenced by our unconscious registering of relevant factors, a registering that never makes it through to the level of judgement.[15]

So to summarize these considerations: I think that part of the reason why people are convinced they have freedom of the will is that they have an experience of choice, and that this experience corresponds to a real phenomenon: a conscious process of forming an intention to perform a certain action from a range of possible actions, and then, if all goes well, performing that action. But this is a very specific phenomenon. I have said that choice is frequently necessary for action, but it is clear that it is not always so. It is only when the questions of what to do explicitly arises that we need to make a choice. Many of our actions are habitual, or otherwise automatic. It has long been appreciated that certain motor actions, once mastered, require no conscious thought. Indeed, conscious thought can be inimical to them: the movement of one's feet as one runs downstairs is, to take William James' example, best left unconsidered. But much recent work in social psychology has pushed the class of automatic actions much further, to cover much of our routine activity.[16] We can wend our way through a big city, safely using various modes of transport and completing various social interactions, without making any choices. It is only in certain circumstances—when we enter a supermarket, say, or a restaurant, or a cinema with many screens—that choices have to be made.

Clearly we think of many of these automatic actions as in some sense free, but our reason for thinking this cannot be that they are chosen. What is it then? One possibility is that there is a distinctive phenomenology governing them too. Here things are, I think, less than clear. There is not obviously a phenomenology of agency, as we might call it; but there does seem to be something like a phenomenology that occurs when agency is lost, or is perceived to be lost. Patients with anarchic hand syndrome, whilst conceding that the actions of their anarchic hands are in some sense theirs, report that they do not feel to be.[17] Even more strikingly, schizophrenic patients with delusions of alien control have the experience that their actions are under the control of

---

[14] (Holton, 2009) Ch. 3.

[15] For instance, it looks as though we can be sensitive to patterns in the world, so that our choices may be influenced by them, even though we form no conscious beliefs about them. In such cases our choices may be much better than chance, though we have no idea why. I discuss some examples in (Holton 2009) Ch. 3.

[16] See, e.g. (Bargh and Chartrand 1999), (Bargh 2002)

[17] 'Of course I know that I am doing it' says a patient of Marcel's; 'It just doesn't feel like me'. See (Marcel, 2003) p. 79.

others.[18] They can even accept that the have the intentions to do the things that they are doing, whilst denying that those intentions are causing the actions. Plausibly then, as some have suggested, there is some feedback system that produces a distinctive phenomenology in cases in which our intentions are not achieved, and it is this that is mistakenly triggered in the schizophrenic patients.[19] If this is right, and the phenomenology is basically negative rather than positive, it should be possible to generate a conviction in subjects that they are performing actions even when they are not; and there are some reasons for thinking that this can be done.[20]

A phenomenology of loss of agency is something, but since it is somewhat recherché, it seems unlikely that it is the source of our notion. So rather than positing a distinct phenomenology, an alternative is to tie agency back to choice using a dispositional account. Although we do not choose the movements of our feet as we run downstairs, we could; or at least, we could choose which movement to make with each foot at each moment, though the movements would undoubtedly be far from fluid. Likewise with the other automatic actions that we perform. We think of these are free actions exactly because we have the capacity to bring them under the control of choice, which in turn has two aspects: first the capacity to choose, and second the capacity to turn that choice into action.

I speak in terms of capacities here, rather than saying that in such cases we could have acted otherwise. The reasons for doing so are familiar amongst philosophers from cases made famous by Harry Frankfurt.[21] Black has implanted a device in Jones's brain so that, should Jones not respond to his bidding, Black could take control of him. In fact Jones does oblige without Black ever needing to activate the device. Jones, we think, is responsible for what he has done, responsible in a way that he would not have been had the device been used; yet it is not true that he could have acted otherwise. Some have concluded that moral responsibility is independent of the capacity to choose. But this is surely too quick. Instead we should realize that an agent can have the capacity to choose, and can exercise it, even though it is not true that they could have chosen to do otherwise, or that they would have done otherwise if they had so chosen. In short,

---

[18] For example: 'My grandfather hypnotized me and now he moves my foot up and down', 'They inserted a computer in my brain. It makes me turn to the left or right', 'The force moved my lips. I began to speak. The words were made for me', 'It's just as if I were being steered around, by whom or what I don't know'. (Frith, Blakemore and Wolpert, 2000) p. 358.

[19] In contrast, patients with anarchic hand syndrome do not attribute the actions of their anarchic hand to anyone else, nor do they think they have the intentions to perform the actions that their hands are performing. It seems that they are right in this, since, unlike in the case of the schizophrenic patients, a major part of their motor control system is not working. For discussion of the likely neurophysiology here, see (Della Sala, 2005).

[20] See Wegner and Wheatley's I-Spy experiments, (Wegner and Wheatley, 1999). As Eddy Nahmias pointed out to me, the results are far from conclusive.

[21] (Frankfurt 1969)

dispositions should not be analysed in terms of counterfactuals. This is a point that has been long known in the literature on dispositions.[22] For instance: a glass is fragile in virtue of its internal constitution, even though a protecting angel ensures it will not break if dropped. It has the disposition of fragility even though it is not true that it would break if dropped. We can easily modify this into a case that mirrors that of Black more closely. A fair coin is tossed, and comes down, quite without interference, heads. However, were it to have been about to come down tails, an interfering genie would have intervened and flipped it over. Clearly under these conditions the coin couldn't have come down other than heads and so wouldn't have come down other than heads. Equally clearly, where the genie doesn't intervene, it has freely, fairly, fallen heads.[23] The same, I suggest, is true of Black: he acted freely since he exercised his capacity to choose and to act, even though it is not true that he could have acted otherwise. So I confine my talk to capacities.[24]

Let me sum up this section: our experience gives us access to two different capacities, choice and agency. Neither is incompatible with physical determinism, though both are incompatible with psychological determinism (i.e. with determination by conscious beliefs, desires and intentions), which might be mistaken for it. I suggest then that whilst physical determinism may be true (it is up to physics to tell us whether it is), psychological determinism is probably false (small wonder then that I am reluctant to accord much weight to the opinions of Nahmias's subjects who conclude the other way). However, my focus will not be on the relation of these capacities to metaphysical theses about determinism, but on the relation they bear to morals.

Morals for Morals

Can either the capacity for choice, or the capacity for agency, shed much light on the conditions needed for moral responsibility? I start with choice. There is a model of responsible action that I suspect is behind much moral theorizing. The agent investigates the world, determines the possible courses of action, chooses which action to perform, and then performs it as a result of that choice. All is transparent and deliberate. There is no doubt that in such circumstances agents are standardly responsible for their actions (standardly, since for all we have said there may be other conditions that are not fulfilled; we are now looking at choice as a putative necessary condition, and not a sufficient one). So clearly choice can play an important role. But, just as we have seen that choice is not necessary for action in general, so it is surely the case that it is not necessary for those actions for which we are morally responsible. I do

---

[22] See (Martin, 1994); Martin's point had been widely known for at least twenty years before that.

[23] Kadri Vihvelin, from whom I take the example (Vihvelin, 2000), draws the different conclusion that the coin could come down tails. The reasoning, which strikes me as mistaken, seems to be based on reading the 'could' claim as a counterfactual, rather than a simple statement of possibility.

[24] For some similar thoughts see (Fischer, 1994) 154–8.

not just mean that we can be responsible in cases in which we are negligent, cases where, for instance, we harm someone without choosing to do so because the act we do choose to do has their harm as a readily foreseeable consequence. I mean rather than we are frequently held responsible for automatic actions. Suppose I have a good-sized whisky every day before driving home from work, and that over the years the quantity I pour myself has crept up, so that now it takes me well over the legal limit to drive. I may go through the whole process quite automatically, writing memos, talking on the phone, tidying things away as I pour and drink. My moral culpability is undiminished.

So choice is not a plausible necessary condition on morally responsible action. What about agency? This is rather more promising. If I kick you because someone trips my knee jerk response, or because I am in the throes of epileptic attack, I am not held to blame.[25] In both cases I lack the capacity to choose to do otherwise. More contentiously, if I promise, but fail, to give up the cigarettes to which I am, unknowingly, addicted, many would hold that I am not to blame, since whilst I have the capacity to form the intention to give up, I lack the capacity to carry it out.[26]

So there are good grounds for thinking that agency is a necessary condition on moral responsibility. However, it is rather a weak condition. I doubt that it is what is wanted when people say that to be morally responsible one must be free. When we formulate necessary conditions we want them to be as restrictive, and hence as informative, as possible. This one is not sufficiently restrictive. Consider a case of paranoid schizophrenia: the sufferer, convinced that he is about to be attacked, hits out at someone who is in fact, and quite transparently, innocent. We acquit him of blame just as we acquitted the epileptic. Yet he was clearly exercising his agency as we have described it: he had the capacity to choose what to do and he acted on that choice. His defence comes from the fact that he is deluded in the beliefs on which he acts.

If we want to capture this failing as a failure of agency, we will need a more restrictive notion. The free agent has not just a capacity to choose, but a capacity to choose *rationally*; that is what the person suffering from paranoid schizophrenia lacks. This is the kind of condition for which Michael Smith has recently argued; he sees it as capturing the 'kernel of truth' in the doctrine that moral responsibility requires the ability to do otherwise.[27] Smith tries to develop the account in terms of a set of counterfactuals: to have the capacity to act rationally requires that one would act rationally in a set of counterfactual conditions. We have seen good reason to doubt that capacities can be reduced to single counterfactuals; I am unsure that they can be reduced even to sets of counterfactuals. So let us leave that issue aside, and focus

---

[25] In English law these are covered by the defence of automatism.

[26] I take it that at least one source of the contentiousness of these cases reflects the contentiousness of the latter claim. But I suspect that these may be cases in which we would blame even if we thought that the agent lacked the capacity to resist. See below.

[27] (Smith, 2003).

instead on the simple claim that to moral responsibility requires the capacity for rational choice.

The thought is something like this: to be morally responsible for an action, one must have the capacity to rationally assess the reasons for or against doing it, and to form and act on one's intentions accordingly; or, in Smith's words, agents are only morally responsible for 'those things that happen as a consequence of their responding or failing to respond to reasons to the extent that they have the capacity to do so.'[28]

There is certainly something intuitively appealing about this view. And whilst it doesn't quite coincide with the conception of freedom that we developed in discussing agency, it is certainly a natural extension of it: we simply add a rationality requirement to the capacity for action. Nevertheless, I am inclined to think that it cannot be right. It is too restrictive as a condition on those we hold morally responsible. Our moral practice involves us in criticizing agents who do not meet it. The conception that Smith develops means that moral requirements are tailored very closely to the actual capacity of the individual concerned: and that is not something that we are prepared to do.

To see what this might mean, consider an alternative conception of moral requirements that has been well expressed by Pamela Hieronymi.[29] On this alternative, requirements are relatively insensitive to the abilities of the agent, in a way that makes them more like other demands. As she says, 'in many areas of adult life—in one's career, in one's role as teacher or parent, in one's position as chair or as second tenor—the demands one is under remain insensitive to one's own particular shortcomings; one's capacities develop as one tries to meet them.'[30] Likewise for moral demands: they can impose upon us even if we lack the capacity to meet them.

Hieronymi gives a number of considerations in support of this conception, starting from the aspirational nature of our moral demands, and from the character of blame. I shall take a rather different course, exploring some considerations in favour of the conception from our ordinary moral practice.

Let us return to the Calvinist theme raised earlier, and consider the writings of Daniel Dyke, a English Puritan writing in the early Seventeenth Century. Amongst many works, Dyke wrote a tract entitled *The Mystery of Selfe-Deceiving*, whose contents are, to a contemporary ear, more aptly revealed by its subtitle: *A Discourse and Discovery of the Deceitfulnesse of Mans Heart*. The self-deception of which Dyke is concerned is simply self-ignorance, in particular, ignorance of one's own motives. 'God only knoweth the heart *exactly* and *certainly*: Because man and Angels may know it conjecturally, and by way of guessing'.[31] Insofar as we can have self-knowledge, it is only

---

[28] (Smith, 2007) p. 142

[29] (Hieronymi, 2007); see also (Hieronymi , 2004).

[30] (Smith, 2003) p. 111

[31] (Dyke, 1630) p. 399

if we have 'plowed with Gods Heifer',[32] but even then the knowledge is partial. 'Onely God of himselfe exactly knoweth the secrets of the heart. There is a great mingle-mangle and confusion of thoughts, even as there is of drosse and good metall in silver and gold, which lie so confused together, that to the eye of man the drosse in not discernable.'[33]

As a Puritan, Dyke was, as I have said, a Calvinist, and clearly the doctrine of predestination is central to the views expressed here. Salvation comes entirely from God's grace, grace that is extended only to the elect, and so none but God one can know for sure who is destined for it. Similarly, at the level of the individual action, we cannot know what is truly pleasing to God, since we cannot know what is done from the right motive.[34] Dyke is thus rejecting the kind of account that Smith offers. We do not need to be able know our motives to behave righteously; good action does not require the kind of rational capacity on which Smith insists.

Dyke's was writing in the Seventeenth Century, but the views about self-knowledge have a strikingly contemporary ring. I suggest that we should draw the same conclusions for ethics that he draws. It has become a commonplace that our motives are not transparent. Self-deception, especially about motives, is the standard condition for us to be in. Indeed, a wide range of studies have concluded that anything approaching self-knowledge is, in many areas, had only by the depressed.[35] But, if we cannot know our own motives, how can we go in for the kind of rational assessment of our own actions that an account like Smith's requires?

Let us take an example, one that is not too grandiose. Suppose that Emma is deeply interested in her friend's romantic attachments, offering advice and persuasion that, given her superior standing, is sure to be taken up.[36] Suppose further that Emma thinks that she is acting entirely for the welfare of her friend, but she is wrong: putting her many actions together, the insightful and disinterested observer can see that she is moved by a certain vanity: by a desire to create and control those she thinks beneath her, and to do so, moreover, in ways that will bring certain advantages to her. Suppose finally that she would be horrified were she to realize that this is the case, and would immediately stop. She is not malicious; just self-deceived.

---

[32] *Ibid*. The reference is to *Judges* 15.18; his opponents having pressured his wife into revealing the answer to a riddle, Samson replies 'If ye had not ploughed with my heifer, ye had not found out my riddle'. This is particularly interesting, since the idea here is surely that pressuring his wife is, in one commentator's words, 'an unworthy expedient' (Jamieson, Fausset and Brown, 1961). Could Dyke think that self-knowledge is not proper to man?

[33] *Ibid*. p. 402

[34] Moreover, God's knowledge of the heart of man is the kind of maker's knowledge that assumes predestination: God created man ('Artificers know the nature and properties of their works; and shall God onely be ignorant of his workmanship?' p.403), and is also 'the preserver and upholder … of the motions of the mind' *ibid*. I discuss some of these themes further in (Holton 2000)

[35] Starting with (Alloy and Abramson, 1979)

[36] For a rather fuller picture, see Jane Austen, *Emma*.

This is a moral failing in Emma, no doubt. But do we make that judgment because we are confident that she has the ability to see her error? Would we withdraw the verdict if we discovered that she were simply incapable of doing so? I think not. Of course, if she were so lacking in insight that the case were pathological, then we might withdraw all moral censure, and treat her as a patient. But if she simply has a standard amount of self-deception, and it turns out that there was no way to move her from it, then that would provide no excuse. Her self-ignorance may not be itself a moral failing, but it can lead her to moral failure.

The point applies quite generally. For many moral failings, culpability requires the relevant bad motive. But we do not require that in addition the agent have, or be able to have, *knowledge* of that motive. A person can be spiteful, or selfish, or impatient without knowing, or being able to know that they are; and such ignorance does not excuse the fault. Likewise, for criminal offences, the law sets down the appropriate *mens rea*. This will typically require knowledge of certain facts about the external world—that the action was harmful, or likely to be harmful, that object taken was the property of another, and so on—but in it never requires knowledge of the very state of mind. It is no legal defence that one is ignorant of one's own motives. People can be guilty whilst honestly believing themselves to be innocent, just as they can be innocent whilst believing they are guilty.

These considerations are relevant to another issue, and it may be fruitful to spend a little time pursuing it. In a series of recent articles, Gideon Rosen has argued that mistake provides a quite general moral defence.[37] Incontestably, one cannot be guilty of lying if one thought one was telling the truth, and one cannot be guilty of malice if one thought one was acting in the victim's best interests. But Rosen wants to push the idea further. It is not just ignorance of the facts that can provide a moral defence, but also ignorance of ethical principles: if someone non-culpably believes that what they are doing is right, then that too is a defence. The upshot, Rosen concludes, is a form of moral scepticism: since we cannot be sure that people were acting contrary to their moral beliefs, we cannot be sure that they are morally responsible.

To me this has the air of *reductio*. Rosen is pushing a piece of moral theory against what are often called Moorean facts: facts about which we are more certain than we are of any bit of philosophy that might seek to overturn them.[38] Many of the most monstrous crimes of the Twentieth Century have been perpetrated by those who appear to have thought that what they were doing was right—Pol Pot, for instance, or many of the leaders of Nazi Germany. The belief, even if non-culpably held, that the killing of one's political opponents is morally justifiable does nothing to excuse the action.

---

[37] (Rosen, 2003, 2004)

[38] The reference is to G.E. Moore's argument for the existence of the external world, an argument that took it as a premise that he had hands, something he took as more certain then the philosophical premises that sought to undermine it. See (Moore 1939).

The common law takes a similar perspective. The maxim that ignorance of the law is no defence, whilst it does not hold absolutely, is nevertheless central to Anglo-American jurisdictions.[39] It is sometimes seen as driven purely by expediency, a device to discourage wilful blindness. I think, however, that to insist that this is all there is to the doctrine is to miss much of what is central. For the thought is that the core demands of the common law are simply binding upon everyone in the society, even if, for some reason, they are ignorant of those demands. Of course things are different if the agent is truly mad. Most jurisdictions have an insanity defence, typically along the lines given in the Model Penal Code: a person is not responsible if 'as a result of mental disease or defect he lacks substantial capacity either to appreciate the criminality [wrongfulness] of his conduct or to conform his conduct to the requirements of law'.[40] But a 'mental disease or defect' as the courts have interpreted, it requires much more than simply the inability to know what law or morality demands. It requires a whole set of distortions that affect much of what the subject does. Indeed, the Model Penal Code explicitly holds that the insanity defence does not extend to 'an abnormality manifested only by repeated criminal or otherwise antisocial conduct'.[41]

So both our ordinary moral practices and the law reject the idea that responsibility requires the rational capacity to act well. In criticizing Emma we do not need to know whether she really had the rational capacity to realize that what he was doing was wrong. The same point, though on a totally different scale, applies when we criticize Pol Pot. In saying this though, I do not mean to suggest that we feel no conflict in these cases. What happens is that we have a clash between a certain theoretical view—responsibility requires the ability to behave otherwise—to which we are well wedded; and our views about particular cases—this person is responsible—to which we are even more closely wedded. That should remind us of one of the findings mentioned earlier: that, when they think in the abstract, people tend to judge determinism incompatible with moral responsibility, but that when they consider concrete cases they do not. The only way out is to change the requirement expressed in the theoretical

---

[39] See the Model Penal Code 2.04 for a codification of when ignorance of law can provide a defence (American Law Institute, 1985). With a few exceptions, it is limited to cases in which it undermines the *mens rea* for the offence. For discussion of the similar doctrine in English law see (Smith, 1999) pp. 82–4.

[40] (American Law Institute, 1985) 4.01. This provision is based on the English M'Naghten Rules. The U.S. Insanity Defense Reform Act of 1984 required in addition that the mental disease or defect be *severe*.

[41] Here I part company with several philosophers, in particular with Susan Wolf, who argues in an influential piece that the inability to form moral judgements accurately does preclude moral criticism, and is to that extent a form of insanity (Wolf, 1988). Wolf bases her argument around the M'Naghten rules, though she reads them in a very different way to the way in which a court would read them. In particular, her central example—JoJo, the son of a brutal dictator whose spoilt upbringing leaves him morally incompetent—is not someone who we should normally think of as lacking in legal or moral responsibility. Or at least, the closest real-life examples we have—Jean-Claude 'Bébé Doc' Duvalier of Haiti comes to mind—are not people typically judged to lack such responsibility.

view; and a first step in doing that is to realize that we are not dealing with a single notion of freedom.

These are rather grand themes, and I have treated them all too briefly. But I hope I have done enough to sketch the form of a position according to which, in so far as we have a notion of a free act that provides a condition on moral responsibility, it is not that of either choice or agency. I haven't said very much about what it is. My own view is that getting clearer on it will require getting a lot clearer on how our emotional and rational faculties work when making moral judgements. And, despite much recent work, that is a topic about which we still know rather little.

Bibliography

Alloy, Lauren and Lyn Abramson, 1979: 'Judgment of Contingency in Depressed and Nondepressed Students: Sadder but Wiser?', *Journal of Experimental Psychology: General* 108, pp. 441-85.

American Law Institute, 1985: *Model Penal Code* (Philadelphia: The American Law Institute)

Bargh, John: 2002: 'Losing Consciousness', *Journal of Consumer Research* 29 pp. 280–5.

— and Tanya Chartrand 1999: 'The Unbearable Automaticity of Being' *American Psychologist* 54 pp. 462–79.

Baumeister, Roy, E. Masicampo and C. Nathan DeWall, Forthcoming: 'Prosocial Benefits of Feeling Free: Manipulating Disbelief in Free Will Increases Aggression and Reduces Helpfulness'.

Della Sala, Sergio, 2005: 'The Anarchic Hand' *Psychologist* 18, 606–609.

Dyke, *The Mystery of Selfe-Deceiving* (Revised Edition, London: Richard Higgenbothan, 1630)

Fischerr, John Martin 1994: *The Metaphysics of Free Will* (Oxford: Basil Blackwell)

Frankfurt, Harry 1969: 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy*, 66, pp. 829-839

Frith, Christopher, Sarah-Jayne Blakemore and Daniel Wolpert 2000: 'Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action'. *Brain Research Reviews*, 31, 357–363.

Hieronymi, Pamela 2004: 'The Force and Fairness of Blame', *Philosophical Perspectives* 18, 115–48

— 2007: 'Rational Capacity as a Condition on Blame', *Philosophical Books* 48, 109–23

Holton, Richard 2000: 'What is the Role of the Self in Self-Deception?' *Proceedings of the Aristotelian Society*, 101, pp. 53-69

— 2009: *Willing, Wanting, Waiting* (Oxford: Clarendon Press)

Jamieson, Robert, A. R. Fausset and David Brown 1961: *Commentary on the Whole Bible* (Grand Rapids: Zondervan).

Kripke, Saul 1972: *Naming and Necessity* (Cambridge: Harvard University Press)

Marcel, Anthony 2003: 'The Sense of Agency' in J. Roessler and N. Eilan (eds.) *Agency and Self-Awareness* (Oxford: Oxford University Press) 48–93,

Martin, Charlie, 1994: 'Dispositions and Conditionals', *Philosophical Quarterly* 44

Nahmias, Eddy, Stephen Morris, Thomas Nadelhoffer, and Jason Turner, 2005: 'Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility,' *Philosophical Psychology*, 18, pp. 561–84.

Nahmias, Eddy, D. Justin Coates, and Trevor Kvaran, 2007: 'Free Will, Moral Responsibility, and Mechanism: Experiments on Folk Intuitions' *Midwest Studies in Philosophy* 31 pp. 214–42.

Nichols, Shaun and Joshua Knobe, 2007: 'Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions', *Noûs*, 41 pp. 663-685.

Nichols, Shaun and Adina Roskies, forthcoming: 'Bringing Moral Responsibility Down to Earth', *Journal of Philosophy*.

Nietzsche, Friedrich, 1886: *Beyond Good and Evil*. Trans. R. Hollingdale (Harmondsworth: Penguin, 1973).

Putnam, Hilary (1975): 'The Meaning of "Meaning.' *Philosophical Papers, Vol. II : Mind, Language, and Reality*, Cambridge: Cambridge University Press.

Quine, W.V.O. 1951: ' 'Two Dogmas of Empiricism,' *The Philosophical Review* 60, 20–43

Rosen, Gideon 2003: 'Culpability and Ignorance' *Proceedings of the Aristotelian Society* 103 61–84;

— 2004: 'Skepticism about Moral Responsibility' *Philosophical Perspectives* 18, 295–313.

Smith, John, 1999: *Smith and Hogan's Criminal Law* (London: Butterworths)

Smith, Michael, 2003: 'Rational Capacities', in Sarah Stroud and Christine Tappolet (eds.) *Weakness of Will and Practical Irrationality* (Oxford: Oxford University Press) .

— 2007: Reply to Hieronymi in 'In Defense of *Ethics and the A Priori*' *Philosophical Books* 48 p. 142.

Sorabji, Richard 2000: *Emotion and Peace of Mind: From Stoic Agitation to Christian Temptation* (Oxford: Clarendon Press).

Vihvelin, Kadri, 2000: 'Freedom, Foreknowledge, and the Principle of Alternate Possibility', *Canadian Journal of Philosophy* 30, 1–23.

Vohs, Kathleen, and Jonathan Schooler, 2008: 'The Value of Believing in Free Will' Psychological Science, 19, pp. 49–54.

Wegner, Daniel, 2002: *The Illusion of Conscious Will* (Cambridge: Harvard University Press).

— and Wheatley 1999: 'Apparent Mental Causation: Sources of the Experience of Will' *American Psychologist* 54, 480–92.

Wolf, Susan 1988: 'Sanity and the metaphysics of responsibility' in Ferdinand Schoeman (ed.) *Responsibility, Character and the Emotions* (Cambridge: Cambridge University Press) pp. 46–62.