

## HOLD OFF ON THE DEFINITIONS: COMMENTS ON BAUMEISTER

RICHARD HOLTON, MIT

Roy Baumeister's paper bristles with ideas. Rather than plunging into the details, let me survey the landscape. From a philosopher's perspective, the paper falls into two parts. The first, much shorter, reacts to some skeptical philosophical arguments that seek to question the existence of free will. With these out of the way, the second part attempts to build a scientific theory of free will. Or, if that sounds a little too grand, it attempts to assemble some of the main parts of the parts from which such a theory will be built.

I think that first part, the response to the philosophical arguments, doesn't quite work. Fortunately however, I think that this doesn't matter for the second, positive, part. Indeed I think that the work that Baumeister outlines there is exactly the kind of thing that is needed for any theory of free will.

So my aim here is to return to the first part, and to do a little philosophical under-laboring in support of Baumeister's positive picture. It is not that I have a better response to the skeptical philosophical worries that he sketches, or at least, not directly. It is rather that I think that they better side-stepped than approached head-on. A useful comparison here is with epistemology—the theory of knowledge. Twentieth century epistemology was dominated by two themes: (i) an attempt to give a definition of knowledge, and (ii) an attempt to answer Cartesian skepticism, the worry that we might be massively mistaken in our views (that we might be deceived by a demon, or, in its more contemporary form, that we might be brains in vats). The two themes are of course linked: the traditional route is to provide a definition of knowledge, and then show how it answers skepticism. If Twentieth century epistemology has shown anything, it has shown that such a route is hopeless. Finding counter-examples to definitions became the philosopher's standard move, and learning to find them became part of the basic training of every graduate student. Almost always they were successful. Definitions of substantial philosophical notions hardly ever work. Seen with a little distance, this should not be surprising. Why would we expect that one notion can be defined in terms of others?

In recent years progress in epistemology has been made by taking a quite different tack, by starting with more mundane cases of knowledge and exploring what is distinctive about them, without trying for a definition or a response to skepticism. Perhaps in time a response to skepticism will come out of such work, though I doubt that it will ever satisfy the serious skeptic. This should not, however, be too concerning. The need is not to satisfy the skeptic; it is rather to show why those of who do not start out as skeptics (which in reality is pretty much all of us) can get clearer on the phenomena without being driven into skepticism.

I suggest that much the same has happened with philosophical accounts of free will. There the agenda has been set by the skeptic: How can we have free-will if determinism is true? And the typical response, starting with Hobbes, has been to start by trying to find a definition of free will that can be shown to be compatible with determinism. Characteristically such definitions don't work: either they fail to capture some element of free will, or they fail to show how they are compatible with

determinism.<sup>1</sup> Here too I suspect that the problem comes from the attempt to work from a definition of free will; and the right response is to start elsewhere. So let me explore this in a little more detail, starting with Baumeister's own arguments against the skeptic.

Baumeister reviews three skeptical arguments. The first, the idea that free will must be exempt from causation, he dismisses. I join him in that. The interest then is in the second and third. Both of these are deterministic arguments of a kind, but one is to do with determinism of one level by another, whereas the other is to do with temporal determinism. I take these in turn.

The issue of levels has had a good degree of discussion in recent years; indeed there has been a wave of concern in the mainstream press provoked by the idea that the psychological is determined by the neurological. "Recent science has shown that all of our thought processes are just firings of neurons" goes the worry, "but neurons can't be free, so there is no free will." And if that is not worrying enough, the argument can be taken down a level further: "Firings of neurons are just interactions between subatomic particles, and subatomic particles certainly can't be free". Baumeister objects that this argument requires a reductive premise: the idea that explanations at one level can be reduced to explanations at a lower level. And that he thinks is implausible. Different levels have different patterns of explanation. Explanations at the psychological level cannot be reduced to explanations at the physical level; all of science will not be reduced to fundamental physics.

The non-reductive line that Baumeister proposes here is well-represented in philosophy.<sup>2</sup> But even if we accept it, it is not clear that it really addresses the heart of the worry. For concerns about levels have never on their own been central to philosophical worries about free will. After all, to take a very rough analogy, the fact that tables are made out of atoms doesn't, on the face of it, show that there are no such things as tables (though of course a few philosophers have disagreed). Similarly then, the idea that all mental activity is somehow grounded in the behaviour of neurons, and that that in turn is grounded in the behaviour of subatomic particles, doesn't by itself challenge the idea that we have free will. Rather, these claims have been important in so far as they interact with ideas of the other kind of determination, namely temporal determination, the idea that how things stand at one time is determined by how they stood at an earlier time. And all that one needs to get these worries going is not reductivism but the much weaker thesis of supervenience.

The idea of supervenience is simple enough (although it can be complicated to state precisely). Consider a picture of the Taj Mahal on a computer screen. The picture is made up of an arrangement of pixels. Switch off the pixels and you lose the picture. If someone looked carefully into the screen wanting to find the picture, and complained that all they could find was the pixels, you'd explain that they had misunderstood: in some important sense there is nothing to the picture but the pixels. But in another sense the picture doesn't reduce to the pixels. We could doubtless come up with a qualitatively indistinguishable picture which used a different array of pixels (we are not so sensitive to the pixels to pick up on any difference). And, once we think more abstractly, it is clear that there are a huge number of arrangements of pixels which, whilst very different to this one, would still give rise to pictures

---

<sup>1</sup> For a very persuasive case along these lines see W. Lycan, 'Free Will and the Burden of Proof' in Anthony O'Hear (ed.), *Minds and Persons* (Cambridge: Cambridge University Press, 2003).

<sup>2</sup> See for instance, Stephen Yablo 'Mental Causation' *The Philosophical Review* 101 (1992), pp. 245–280 for a classic defense.

of the Taj Mahal. In the standard terminology, such pictures can be *multiply realized* at the level of pixels. Whilst such pictures will all be pictures of the Taj Mahal, there may be precious little that they have in common at the level of pixels. So there will be many things that we might want to say about them that cannot straightforwardly be said by talking about pixels.

It is here that the terminology of supervenience comes in. We say that this picture of the Taj Mahal supervenes on the arrangement of the pixels, even though it doesn't reduce to it. The central idea is that there could be no change in the picture without a change in the pixels. Crucially though, in cases like this, the relation is not symmetric. We could change the pixels and still have a picture of the Taj Mahal. So the arrangement of pixels does not supervene on the picture.

Supervenience provides a neat way of elaborating a kind of non-reductive physicalism that should appeal to Baumeister.<sup>3</sup> The idea is that the mental supervenes on the physical. Such an account is physicalist in that we accept that the building blocks of the mental are the neurons, and in turn the smaller particles that make them up; someone who looked into the brain and objected that they could find no ideas there, only neurons, would be badly off track.<sup>4</sup> But it is non-reductive in that the multiple realizability of the mental means that mental explanations cannot be usefully couched in terms of neurons.

The next step is to see that this comparatively weak thesis of supervenience, when coupled with a thesis of temporal determinism at the level of the physical, is enough to generate the skeptical challenge to free will. Temporal determinism at the level of the physical says that, given how things stand at one time,  $t_1$ , and given the laws that obtain, it is determined how things will stand at a subsequent time,  $t_2$ .<sup>5</sup> If the mental supervenes on the physical, then an individual's mental state at any time will supervene on the physical state of things at that time. But since their mental state at  $t_2$  will supervene on the physical state at  $t_2$ , and since the physical state at  $t_2$  is determined by the physical state at  $t_1$ , then there is an important sense in which their mental state at  $t_2$  will be determined by the physical state at  $t_1$ . In short, supervenience requires that the mental cannot change without the physical changing; but if the physical is fixed by how things were before, then the mental is fixed by how things were before. And to view the mental as fixed, runs the worry, is to deny the possibility of free will. For after all, the mental includes the decisions that one makes; and if there is only one decision that one can make, how can one have free will?

---

<sup>3</sup> Baumeister mentions supervenience towards the end of his piece, but I don't think he quite sees its relevance for the sort of position he wants to embrace.

<sup>4</sup> Baumeister does at one point say that "shared understandings are not made of molecules." So perhaps he would reject even the supervenience claim. But if they are not in some sense made up of molecules, what are they made of? I take it that Baumeister's real intent here is to say, as he says in the preceding sentences, that they will not be *explained* in terms of a theory of molecules.

<sup>5</sup> Baumeister at times puts this in terms of knowledge: knowing the laws and initial conditions enables us to know how things will evolve. But this is a much stronger claim, which I think is quite implausible for creatures like us who can act to frustrate any prediction. (For discussion see my ). As a scholarly point: Baumeister imputes to Calvin the idea that predestination stems from God's knowledge of how things will happen. Whilst there was a substantial medieval concern with God's foreknowledge, Calvin's primary concern, like that of other Reformation thinkers, was with God's *power*. We are predestined for salvation or damnation because God has either extended his grace to us or he has not; to deny this would be to place a limit on his power.

I am not quite sure what Baumeister's response to this worry is. He plays with the idea that the physical is not determined, but is rather probabilistic. That itself is a much debated issue within physics and the philosophy of physics, and it would be, to say the least, worrying if the existence of free will were to turn on the outcome of that debate. Moreover, a point often made, it is unclear that probabilistic accounts will help, in so far as they offer mere chance in place of determinism. If the physical state at  $t_1$  entails that at  $t_2$  there is a 40% chance that physical state  $x$  will obtain, and a 60% chance that  $y$  will obtain, and so correspondingly a 40% chance that I will be in mental state  $m_x$ , and a 60% chance that I will be in mental state  $m_y$ , how does that make me any freer? In so far as I cannot do anything to control which of those states I go into, that seems to make things worse.

Baumeister's other response is simply to turn his back on the issue. We *have* to think that determinism is false, since it would be "utterly useless as a basis for either psychological theory or for living one's actual life, even if it were true."<sup>6</sup> Surely something has gone wrong here. If determinism is true, it's true, however inconvenient that may be. Perhaps we could take its pragmatic uselessness as *evidence* that determinism is not true. But even that is a bit of a stretch. History is littered with false claims that have made life worth living for many, or that have made certain sciences (or pseudo-sciences) possible.

Nevertheless I think that Baumeister's instinct to turn his back on the skeptical challenge is correct. That is not to say though that we should deny determinism. Rather it is to reject the skepticism that many have thought follows from it. Of course to say this is to accept a form of compatibilism. But my suggestion is not that we should try to provide a definition of free will to show that it is compatible with determinism.

Rather, it is the skeptic who is relying on a definition. How did the skeptical argument work? It moved from determinism to the idea that things could not be otherwise, to the idea that we do not have free will. But that involves two analytic moves: that determinism shows that things could not be otherwise, and that, in the very same sense of 'could', free will requires that they could. Why should we accept those claims? Either because we do some analysis and reason that this is what the notions mean. Or, in slightly more empirical vein, because we do a bit of work and see what most people take the notions to mean. Such an approach can seem inescapable. If we don't start by defining our terms, who knows where we are? We risk just talking past each other.

But we do not have to start with definitions. Chemists didn't start with definitions of the substances they were working with; biologists didn't start with definitions of the biological kinds. Or at least, if they did, they didn't keep with them. Successful definitions came later: they are, at least in part, *discoveries*, findings that come from working with the things themselves, with the substances or the creatures. It is just as well that chemists didn't bind themselves to the idea of an atom as something that is, by definition, indivisible, or that biologists weren't fixed on the idea that mammals must, by definition, give birth to live young.

I think that much the same approach should be taken to free will. If we start with a definition, whether one developed by philosophers or one derived from surveying the masses, we start with something that reflects an implicit prior theory. We shouldn't ignore such theories, since they might embody important

---

<sup>6</sup> Admittedly here he is again talking of Laplacian determinism, i.e. the deterministic thesis phrased in terms of knowledge. But I take it that the idea is also meant to apply to the metaphysical thesis.

insights, but neither should we privilege them. They are prone to embody a vision of how we would like ourselves to be, and such a vision may little encumbered by the facts of how we are. When thinking about epistemology this brings a conception of knowledge as utterly secure, immune from the contingencies of luck; that is what gives rise to Cartesian skepticism. When thinking about free will it brings a conception of free will as immune from the causal forces of determinism; that, I am suggesting, is what gives rise to skepticism.

So, like the chemists and the biologists, we should, as much as possible, start from the other end, from the phenomena of free will themselves, not with our prior views about what they must be. Even the apparently uncontentious definition that Baumeister takes on board—‘Free action means that the person could do different things in the same situation’—needs to be established rather than assumed. And if it is established, this will involve getting clear on just what the ‘could’ means here, which is certainly not something to be decided in advance. As Baumeister says, scientists should be used to the idea that initial ideas will be mostly wrong.

What are the phenomena of free will? Since it is a complex notion, answering to different concerns, there is no simple answer; indeed, part of what we need to discover is whether there is one thing that can play some many different roles. I see at least three. One focus comes from moral concerns: free actions are those for which we are morally responsible. Another comes from some notion of autonomy: free actions are those that we can attribute to someone as really theirs. A third comes from the experience of free will: free actions are those that characteristically give rise to our experience of ourselves as free.<sup>7</sup>

Baumeister’s work tells most immediately on this third dimension. Whilst many philosophers have stressed the experience of freedom, few have stopped to consider what this experience could be an experience of. (It surely isn’t the experience of being an uncaused cause; what would that be like?) Taking willpower seriously, as Baumeister does, provides a plausible account of at least one source of this experience. The central idea here, as I see it, is that choice is a real phenomenon. When faced with an issue of what to do, human beings have the ability to make a choice that is not immediately dictated by their prior beliefs, desires and intentions. Making such a choice is effortful; and in cases where the choice goes against an option that is tempting or otherwise compelling, sticking with the choice is effortful. Moreover, such choices are genuinely effective in determining action—not always, but often enough that the phenomenology of choice is revelatory of a real phenomenon.<sup>8</sup>

As I say, this is what I take to be core. It may be, as Baumeister goes on to say, that such a phenomenon could only evolve in a cultural animal; and it may be that the choices involved are typically concerned with issues of social meaning. But leaving this aside, we already have enough to see that we have many of the characteristic features of free will. And, so far as I can see, such an account is quite compatible with physical determinism. Determinism does not deny that the choices one makes will be effective—denying that is to confuse determinism with a fatalism that actually denies causal efficacy. Further, it is quite compatible with determinism to insist the agent’s beliefs, desires and intentions will not

---

<sup>7</sup> I develop this a little further in ‘Disentangling the Will’ in Al Mele, Kathleen Vohs & Roy Baumeister (eds.), *Free Will and Consciousness: How Might They Work?* (New York: Oxford University Press, 2010). I don’t mean these to be exhaustive.

<sup>8</sup> For more discussion of this, see B. Keith Payne and C. Daryl Cameron ‘Free will worth having and the intentional control of behavior’, this volume.

determine what they will do, for what they do will also be influenced by what they choose, and by how hard they work to put that choice into effect. Of course if determinism is true, then these things will themselves be determined at the physical level; but I find it hard to see how that would bear on our experience one way or the other.<sup>9</sup> At the moment though such idea must remain at the level of first thoughts, awaiting the outcome of more of the sort of experiments that Baumeister and his colleagues have been running. Is this to put philosophers out of a job? I think not. Conceptual work will remain pressing as the results come in. The point is rather that the philosophy is not something that can be got out of the way first. I hope for a long and enlightening collaboration.

---

<sup>9</sup> I develop such ideas further in *Willing, Wanting, Waiting* (Oxford: Oxford University Press, 2009) especially chapters 3 and 8.