

PHILOSOPHY BITES AGAIN

DAVID EDMONDS &
NIGEL WARBURTON



OXFORD
UNIVERSITY PRESS

FREE WILL, RESPONSIBILITY, AND PUNISHMENT

13

DANIEL DENNETT ON Free Will Worth Wanting

David Edmonds: *One way to exercise my freedom would be to act unpredictably, perhaps not to have a typical introduction to a Philosophy Bites interview, or to cut it abruptly short mid-sentence. That's the view of the famous philosopher and cognitive scientist, Daniel Dennett. He also believes that humans can have free will, even if the world is determinist, in other words, governed by causal laws, and he . . .*

Nigel Warburton: *The topic we're focusing on is 'Free Will Worth Wanting'. That seems a strange way in to free will. Usually, the free will debate is over whether we have free will, not whether we want it, or whether it's worth wanting. How did you come at it from this point of view?*

Daniel Dennett: I came to realize that many of the issues that philosophers love to talk about in the free will debates were irrelevant to anything important. There's a bait-and-switch that goes on. I don't think any topic is more anxiety provoking, or more genuinely interesting to everyday people, than free will. But then philosophers replace the interesting issues with technical, metaphysical issues. Who cares? We can define lots of varieties of free will that you can't have, or that are inconsistent with determinism. But so what? The question is, 'Should you regret, or would you regret not having free will?' Yes. Are there many senses of free will? Yes. Philosophers have tended to concentrate on varieties that are perhaps more tractable by their methods, but they're not important.

NW: *The classic description of the problem is this: 'If we can explain every action through a series of causal precedents, there is no space for free will.' What's wrong with that description?*

DD: It's completely wrong. There's plenty of space for free will: determinism and free will are not incompatible at all. The problem is that philosophers have a very simplistic idea of causation. They think that if you give the lowest-level atomic explanation, then you have given a complete account of the causation: that's all the causation there is. In fact, that isn't even causation in an interesting sense.

NW: *How is that simplistic? After all, at the level of billiard balls on a table, one ball hits another one and it causes the second one to move. Neither ball has any choice about whether it moved; their paths were determined physically.*

DD: The problem with that is that it ignores all of the higher-level forms of causation which are just as real and just

as important. Suppose you had a complete atom-by-atom history of every giraffe that ever lived, and every giraffe ancestor that ever lived. You wouldn't have an answer to the question of why they have long necks. There is indeed a causal explanation, but it's lost in those details. You have to go to a different level in order to explain why the giraffe developed its long neck. That's the notion of causation that matters for free will.

NW: *Assuming that you're not going to rely on Aesop here, how did the giraffe get its long neck?*

DD: The lineage of giraffe-like animals gradually got longer necks because those that happened to have slightly longer necks had a fitness advantage over those with shorter necks. That's where the explanation lies. Why is that true? That's still a vexed question. Maybe the best answer is not the obvious one that they got long necks so that they could reach higher leaves. Rather, they evolved long necks because they needed them to drink because they had long legs, and they evolved long legs because they provided a better defence against lions.

NW: *So that's an evolutionary hypothesis about giraffes' necks. How does it shed any light on the free will debate?*

DD: If I want to know why you pulled the trigger, I won't learn that by having an atom-by-atom account of what went on in your brain. I'd have to go to a higher level: I'd have to go to the intentional stance in psychology. Here's a very simple analogy: you've got a hand calculator and you put in a number, and it gives the answer 3.33333E. Why did it do that? Well, if you tap in ten divided by three, and the answer is an infinite continuing decimal, the calculator gives an 'E'.

Now, if you want to understand which cases this will happen to, don't examine each and every individual transistor: use arithmetic. Arithmetic tells you which set of cases will give you an 'E'. Don't think that you can answer that question by electronics. That's the wrong level. The same is true with playing computer chess. Why did the computer move its bishop? Because otherwise its queen would have been captured. That's the level at which you answer that question.

NW: We're often interested in intention where this is linked to moral or legal responsibility. And some cases depend on information that we get about people's brains. For example, there are cases where people had brain lesions that presumably had some causal impact on their criminal behaviour.

DD: I'm so glad you raised that because it perfectly illustrates a deep cognitive illusion that's been fostered in the field for a generation and more. People say, 'Whenever we have a physiological causal account, we don't hold somebody responsible.' Well, might that be because whenever people give a physiological causal account, these are always cases of disability or pathology? You never see a physiological account of somebody getting something right. Supposing we went into Andrew Wiles' brain and got a perfect physiological account of how he proved Fermat's Last Theorem. Would that show that he's not responsible for his proof? Of course not. It's just that we never give causal physiological-level accounts of psychological events when they go right.

NW: I'm still having trouble understanding what an intention is. We usually think of intentions as introspectible mental events that precede actions. That doesn't seem to be quite what you mean by an intention.

DD: When discussing the 'intentional stance', the word 'intention' means something broader than that. It refers to states that have content. Beliefs, desires, and intentions are among the states that have content. To adopt the intentional stance towards a person—it's usually a person, but it could be towards a cat, or even a computer, playing chess—is to adopt the perspective that you're dealing with an agent who has beliefs and desires, and decides what to do, and what intentions to form, on the basis of a rational assessment of those beliefs and desires. It's the stance that dominates Game Theory. When, in the twentieth century, John von Neumann and Oskar Morgenstern invented the theory of games, they pointed out that game theory reflects something fundamental in strategy. Robinson Crusoe on a desert island doesn't need the intentional stance. If there's something in the environment that's like an agent—that you can treat as an agent—this changes the game. You have to start worrying about feedback loops. If you plan activities, you have to think: 'If I do this, this agent might think of doing that in response, and what would be my response to that?' Robinson Crusoe doesn't have to be sneaky and tiptoe around in his garden worrying about what the cabbages will do when they see him coming. But if you've got another agent there, you do.

NW: So, Man Friday appears, and there are problems...

DD: As soon as Man Friday appears, then you need the intentional stance.

NW: So if you have the complexity of interaction that is characteristic of an intentional system, that's sufficient for its having intentions. So there doesn't seem to be any room for the mistake of

anthropomorphism. Anthropomorphism, if the situation is complex enough, is simply the correct attitude to hold towards some inanimate things.

DD: We can treat a tree from the intentional stance, and think about what it needs, and what it wants, and what steps it takes to get what it needs and wants. This works to some degree. Of course, it doesn't have a soul; it's not conscious. But there are certain patterns and reactions. Recently, we've learned that many varieties of trees have a capacity that gives them quasi-colour vision. When the light on them is predominantly reflected from green things they change the proportion of their energy that goes into growing tall. We might say that they have sensed the competition and are taking a reasonable step to deal with the competition. Now, that's a classic example of the intentional stance applied to a tree, for heaven's sake! Fancier versions apply to everything from bacteria, through clams and fish and reptiles and higher animals, all the way to us. We are the paradigm cases.

What's special about us is that we don't just do things for reasons. Trees do things for reasons. But we *represent* the reasons and we *reflect* on them, and the idea of reflecting on reasons and representing reasons and justifying our reasons to each other informs us and governs the intentional stance. We grow up learning to trade reasons with our friends and family. We're then able to direct that perspective at evolutionary history, at artefacts, at trees. And then we see the reasons that aren't represented, but are active. Until you get the level of perspective where you can see reasons, you're not going to see free will. The difference between an organism that has free will and an organism that doesn't has nothing to

do with the atoms: you'll never see it at the atomic level, ever. You have to go to the appropriate design level, and then it sticks out like a sore thumb.

NW: *So we can adopt the intentional stance towards a chess-playing computer, and we probably ought to if we want to beat it at chess, but it doesn't follow from that that it's got free will, or agency?*

DD: Exactly. Those beings with free will are a sub-set of intentional systems. We say 'free as a bird', and birds have a certain sort of free will. But the free will of a bird is nothing compared to our free will, because the bird doesn't have the cognitive system to anticipate and reflect on its anticipations. It doesn't have the same sort of projectable future that we have; nor does it, of course, engage in the business of persuasion. One bird never talks another bird out of doing something. It may threaten it, but it won't talk it out of something.

NW: *So let's go back to the original topic. What is the kind of free will worth wanting?*

DD: It's the kind of free will that gives us the political freedom to move about in a state governed by law and do what we want to do. Not everybody has that freedom. It is a precious commodity. Think about promises. There are many good reasons to make promises: some long-term projects depend on promises, for example. Now, not everybody is equipped to make a promise. Being equipped to make a promise requires a sort of free will, and a sort of free will that is morally important. We can take it apart, we can understand, as an engineer might say, what the 'specs' are for a morally competent agent: you've got to be well informed, have

well-ordered desires, and be movable by reasons. You have to be persuadable and be able to justify your views. And there are a few other abilities that are a little more surprising. You have to be particularly good at detecting the intent of other agents to manipulate you and you have to be able to fend off this manipulation. One thing we require of moral agents is that they are not somebody else's puppet. If you want the buck to stop with you, then you have to protect yourself from other agents who might be trying to control you. In order to fend off manipulation, you should be a little bit unpredictable. So having a poker face is a very big part of being a moral agent. If you can't help but reveal your state to the antique dealer when you walk into the store, then you're going to be taken for a ride, you're going to be manipulated. If you can't help but reveal your beliefs and desires to everybody that comes along, you will be a defective, a disabled agent. In order to maximize getting what you want in life, don't tell people exactly what you want.

NW: That's a very cynical view of human nature! There's an alternative account, surely, in which being open about what you feel allows people to take you for what you really are, not for some kind of avatar of yourself.

DD: Well, yes, there is that. But think about courtship. You see a woman and you fall head over heels in love with her. What's about the worst thing you can do? Run panting up to her showing her that you've fallen head over heels in love. First of all, you'll probably scare her away, or she'll be tempted by your very display of abject adoration to wrap you around her little finger. You don't want that, so you keep something in reserve. Talleyrand once said that God gave men language so

that they could conceal their thoughts from each other. I think that's a deep observation about the role of language in communication. It's essential to the understanding of communication that it's an intentional act, where you decide which aspects of your world you want to inform people about and which you don't.

NW: So freedom, of the important kind, of the kind worth wanting, is freedom from being manipulated. It's about being in control of your life, you choosing to do things, rather than these things being chosen by somebody else?

DD: Yes. In order for us to be self-controllers, to be *autonomous* in a strong sense, we have to make sure that we're not being controlled by others. Now, the environment in general is not an agent, it's not trying to control us. It's only other agents that try to control us. And it's important that we keep them at bay so that we can be autonomous. In order to do that, we have to have the capacity to surprise agents with our somewhat unpredictable trajectory.