

promoting access to White Rose research papers



Universities of Leeds, Sheffield and York
<http://eprints.whiterose.ac.uk/>

This is an author produced version of a paper published in **Philosophical Studies**

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/3313/>

Published paper

Steward, H. (2006) *Determinism and inevitability*, *Philosophical Studies*, Volume 130 (3), 535 - 563.

Determinism and Inevitability

In *Freedom Evolves*¹, Dan Dennett embarks on his second book-length attempt to lay to rest the deep metaphysical concerns that many philosophers have expressed about the possibility of human freedom². One of his main objectives in the earlier chapters of the book is to make determinism appear less threatening to our prospects for free agency than it has sometimes seemed, by attempting to show that a deterministic universe would not necessarily be a universe of which it could truly be said that everything that occurs in it is *inevitable*. In this paper, I want to consider Dennett's striking argument for this conclusion in some detail. I shall begin by suggesting that on its most natural interpretation, the argument is vulnerable to a serious objection. I shall then develop a second interpretation which is more promising than the first, but will argue that without placing more weight on etymological considerations than they can really bear, it can deliver, at best, only a significantly qualified version of the conclusion that Dennett is seeking. However, although I shall be arguing that his central argument fails, it is also part of the purpose of this paper to build on what I regard as some rather insightful and suggestive material which is developed by Dennett in the course of elaborating his views. His own development of these ideas is hampered, so I shall argue, by a framework for thinking about possibility that is too crude to accommodate the immense subtlety and complexity which is exhibited by the workings of the modal verb 'can' and its past tense form, 'could'; and also, I believe, by the mistaken conviction, on Dennett's part, that any naturalistically respectable solution to the problem of free will would have to be of a compatibilist stripe. I shall attempt, in the second half of the paper, to explain what seems to me to be wrong with the framework, and to make some points about the functioning of 'can' and 'could', which I believe any adequate replacement for Dennett's framework must respect. Ironically, though, I shall argue that it is the rejection of Dennett's own framework which holds the key to understanding how to defend the spirit (if not the letter) of his thoughts about the invulnerability of our ordinary modal thinking to alleged threats from determinism.

1. Conway's Life World

It will be important in what follows to note that I shall be working exclusively with Dennett's own definition of determinism, which is borrowed from Van Inwagen (1983)³, and which itself incorporates the modal concept of possibility. Determinism, according to Dennett, is the thesis that "there is at any instant exactly one physically possible future"⁴, not, he says, a particularly difficult idea to understand. But it is nevertheless, he thinks, an idea about which philosophers have been deeply confused. In particular, he claims, philosophers have often supposed that determinism implies inevitability. But according to Dennett, there is no such implication, and the second chapter of *Freedom Evolves* is devoted to explaining why.

Central to the presentation of the main argument of the chapter is the so-called 'Game of Life', invented in the 1960s by the British mathematician John Horton Conway. The Game of Life is played out on a two-dimensional grid of pixels, each of which can be either ON or OFF (thus appearing, for instance, either black or white on a computer screen). Each pixel is surrounded by eight neighbours, as shown below, where "E", the central pixel, is surrounded by eight others:

A	B	C
D	E	F
G	H	I

The state of the grid of pixels changes from one second to the next according to the following simple rule, which Dennett terms ‘Life Physics’:

For each cell in the grid, count how many of its eight neighbours is ON at the present instant. If the answer is exactly two, the cell stays in its present state (ON or OFF) in the next instant. If the answer is exactly three, the cell is ON in the next instant whatever its current state. Under all other conditions, the cell is OFF.

This rule expresses the entire physics of the Life World. Clearly, the Life World is deterministic – given the initial conditions of any given array of pixels (each is either ON or OFF) and the single rule of Life Physics, the future of the pixel array is fixed, assuming, of course, no outside interference with the smooth workings of the system.

Dennett is at pains to stress, however, that we may require an ontology which includes entities of rather greater complexity than the pixel, and accompanying *ceteris paribus* generalisations which exploit a rather higher level of explanation than the single rule of Life Physics, if we are to describe - and in particular, if we are to manipulate - some of the salient patterns which can emerge within the Life World. Gliders, for example, are five pixel configurations which swim across the screen thus:

Insert Diagram: Glider

Then there are eaters – so called because they can destroy or ‘eat’ gliders, and for all I know, other configurations too. Here is the process, described by William Poundstone (quoted by Dennett):

An eater can eat a glider in four generations. Whatever is being consumed, the basic process is the same. A bridge forms between the eater and its prey. In the next generation, the bridge region dies from over-population, taking a bite out of both eater and prey. The eater then repairs itself. The prey usually cannot. If the remainder of the prey dies out as with the glider, the prey is consumed.

Insert Diagram – Eater eating a glider

Such entities – as well as eaters and gliders there are puffer trains, space rakes, loafs, boats, tubs, and others - have a certain robustness within the Life World, and in virtue of their distinctive shapes and capacities for movement, they possess unique powers. By focusing on and harnessing these powers, those aficionados of the Life World whom Dennett calls ‘Life hackers’ are able, by fiddling with the initial configuration of the plane of pixels, to develop interesting creations which are able to persist despite the encroachments of other entities. As Dennett points out, the possibilities for these hackers are, to all intents and purposes, quite endless – a set of Life Worlds just 1 million pixels by 1 million pixels gives you 2 to the trillionth power different possible unfolding universes to explore. Clearly, it would be the height of insanity for a hacker to proceed by means of unguided trial and error through this vast possibility space. Instead, hackers exploit the ontology and the generalisations of the higher level of description, in order to design patterns that will propagate themselves in interesting and beautiful ways over the plane of the Life World. Here are a couple of examples of Life-hacker-speak, quoted by Dennett:

(i) The loaf reacts with all the junk the R-pentomino produces as it naturally transforms into a Herschel, and miraculously reappears some time later leaving no debris at all. It is necessary to prevent the first Herschel glider from hitting the fading remnants of the reaction, and there is no room for an ordinary eater. But luckily a tub with tail and a block can be used instead.

(ii) Dave Buckingham found a faster stable reflector that does not use Paul Callahan’s special reaction. Instead, the incoming glider hits a boat to make a B-heptomino, which is converted into a Herschel and moved round to restore the boat. A compact form of the 119-step Herschel conduit is needed here, as is a non-standard still life to cope with the 64 64 77 conduit sequence.

Dennett’s point in quoting these examples is that hackers work not by thinking about pixels and the single rule of Life Physics, but rather by exploiting the powers of an array of two-dimensional objects which are familiar to them in order to produce the sorts of impressive patterns and systematicities in which they are interested. And it is of especial concern to them how to preserve their creations against the kinds of impacts which might destroy them. There will be much to be said for creations that can avoid particular kinds of harm – which are able, for example, to ‘eat’, move away from or deflect potential attackers, or which can at the very least repair themselves shortly after an encroachment. There is considerable incentive, in other words, to use Life-entities which are good *avoiders* of the various sorts of harm that can befall a denizen of the Life world. And having made this point, Dennett is poised to offer what he says is the main idea of the second chapter of his book - “to show that we need to take the etymology of ‘inevitable’ seriously. It means *unavoidable*”.

Having prepared the ground by means of his introduction to Conway’s Life World, Dennett goes on to offer the following explicit argument in support of his claim that determinism does not imply inevitability:

- P1 In some deterministic worlds, there are avoiders avoiding harms.
- C1 Therefore in some deterministic worlds, some things are avoided.
- P2 Whatever is avoided is avoidable, or evitable.
- C2 Therefore in some deterministic worlds, not everything is inevitable.
- C3 Therefore determinism does not imply inevitability.

In the next section of the paper, I want to consider this argument, and Dennett's response to some of the objections he anticipates. I shall suggest that there is more than one interpretation of the argument, and that the most obvious suggestion to make about how it is supposed to work delivers only an argument with a conclusion which no one – not even the most hardened incompatibilist – would deny. I will then try to suggest an alternative, which however will, in the end, also prove unsatisfactory.

2. Dennett's Argument – Interpretation A

One obvious possibility for someone unconvinced by Dennett's attempt to unyoke determinism from the concept of inevitability by means of reflection on Conway's Life World is to question the first premise of his argument – the idea that the complex creations to which designers in the Life World give rise are real avoiders of anything. In the words of Conrad, Dennett's imaginary interlocutor, "*Real* avoidance involves changing something that was going to happen into something that doesn't happen".⁵ And in Conway's world, of course, nothing that "was going to happen" given the initial condition of the grid and Life physics, is ever avoided. Dennett responds by suggesting that things which were never really 'going to happen' can still perfectly well be avoided; he points out, for example, that a person can avoid being hit by a baseball by dodging it:

But in what sense was that baseball 'going to' hit you smack in the face? You dodged it; you were *caused* to dodge it by the elaborate system evolution has built into you to respond to photons bouncing off incoming missiles on certain trajectories. It was 'never really going to' hit you precisely because it caused your avoidance system to go into action.⁶

There is much that might be said on Conrad's behalf, at this point, and at the end of the paper, I shall suggest that P1 may indeed be problematic, but for now, I intend simply to concede the premise to Dennett. For the purposes of argument, I shall accept that there are things – so as to have a convenient term, I shall call them *fates*⁷ – which are avoided in the Life World. For the main problem with Dennett's argument is not that it has a false premise, but that its conclusions C2 and C3 are too imprecisely stated for it to be clear what exactly each of them means, and that it is not at all obvious how one is supposed to tidy them up in such a way as to generate an argument that is both valid and which has a conclusion that contradicts anything that an incompatibilist is likely to believe anyway.

Let us consider for a moment how C3 is supposed to follow from C2. C2 states that in some deterministic worlds, not everything is inevitable. Given that it is derived directly from C1 and P2, we might take it that this is intended to be equivalent to the claim that in some deterministic worlds some things are evitable, or avoidable. C3 concludes from this that determinism does not imply inevitability. Presumably, what this is supposed to mean is that determinism does not imply that everything is inevitable. But how is the conclusion supposed to be derived? At first glance, it might look as though the intention is to use the existence of deterministic worlds in which some things are evitable as a straightforward counterexample to the suggestion that determinism implies that *everything* is inevitable – the existence of those *some evitable things* contradicting the *everything* of the conclusion. Some of what Dennett offers by way of commentary on the argument suggests that this is indeed what he has in mind, for he notes, in the course of defending his argument, that "pointing to particular instances of avoidance as proof of 'evitability' seems odd",⁸ as though we are indeed to take it that these particular instances are to be regarded as the evitable

things whose existence contradicts the claim that everything is inevitable. But this cannot be satisfactory. For no incompatibilist means to suggest, when she claims that everything is inevitable in the deterministic scenario, that this includes things which *do not in fact happen*. Only what actually happens can conceivably be a candidate for inevitability. She is unlikely to be moved from her conviction that everything is inevitable under determinism by having it pointed out that there are in fact certain things which aren't inevitable, even under determinism – that is to say, various things which don't in fact happen because they were avoided. Another way of putting what is essentially the same point is this: if the argument is to work as an instance of the argument form:

$\exists x$ [x is avoided]

$\exists x$ [x is avoidable (evitable)]

$\neg \forall x \neg$ [x is avoidable (evitable)]

Comment [HS1]: Existential quantifier

Comment [HS2]: Existential quantifier

Comment [HS3]: Negation, Universal quantifier, negation

then the domain over which the variable x ranges will have to include what we might call non-actual fates – things which might have happened, but did not – if the first premise is to be true. But this – even if we allow it, and of course we need not, for quantifying over such possibilities is hardly uncontroversial - only generates a conclusion which is too tame to be worth stating – for of course even the incompatibilist will agree that some of these non-actual fates fail to be inevitable in a deterministic world.

3. Dennett's Argument: Interpretation B

The argument must, then, be more sophisticated than this if it is to work. But what other interpretations are possible? The obvious suggestion to make is that a world in which certain things are avoided cannot be a world in which whatever happens, happens inevitably, because it does not make sense to say that something was avoided unless it could have happened, and what happens cannot be inevitable if other things could have happened in its stead. On this version of the argument, the things avoided are not themselves the evitable things whose existence falsifies the claim that everything is inevitable – it is rather that their evitability is evidence of the possibility of alternative courses of events and their possibility is in turn evidence of the non-inevitability of the *actual* course of events. On this reading, Dennett's summary of his argument represents its structure rather inaccurately, for there are missing premises between P2 and C2 – to the effect that (i) whatever is avoidable could have come about, and that (ii) in situations where alternative fates could have come about, not everything which actually happens is inevitable. But it may nevertheless be reasonable to attribute something like this form of argument to Dennett, particularly if it is permissible to read some of the ideas in his Chapter 3, where Dennett explicitly discusses the concept of possibility, back into the argument of Chapter 2. In the next section, therefore, I shall consider what Dennett has to say about the concept of possibility, and whether it might help us generate a version of his argument which is acceptable. I shall argue, however, that unfortunately for Dennett, the modifications that would be required eventually deliver only a very weak version of the conclusion Dennett is seeking – one with which I think he would be unlikely to be content. Section 5 will show that, in any case, the modifications are problematic.

4. Dennett on Possibility

One of Dennett's main concerns in Chapter 3 of *Freedom Evolves* is to insist that determinism does not reduce our possibilities – that it does not suggest that any of the common sense claims we might be inclined to make about what could have happened – or about what we could have done - is actually false. His strategy is to make a case for the conclusion that in considering the question whether something which did not in fact happen could have happened, we are required to choose a set X of relevant possible worlds with respect to which to consider the question.⁹ Determinism, recall, is the thesis that there is exactly one physically possible future – which means, according to Dennett, that if we consider all the worlds which share a description exactly with our own at any time, and which also share its physical laws, we will find that the set contains only a single world. In this case, we choose the set of worlds, X, by means of what Dennett calls the *narrow method* – that is, we choose the set of physically possible worlds that are *identical* to the actual world at some time t_0 . But he insists that there is no need to choose this method with respect to every possibility question; and indeed that it would be a mistake to do so. He considers, for instance, Austin's famous question whether he could have holed a short putt which in fact he missed.¹⁰ In considering this question, says Dennett, we need to choose the set X of possible worlds which we are to canvass in order to see whether Austin could have holed the putt. If we use the narrow method described above, then we will choose the set X of possible worlds which are identical to the actual world at some time t_0 immediately prior to the putt, and if determinism is true, this set of worlds has only one member, the world in which Austin misses. But we need not choose this method. We may instead admit into the set X a number of worlds which differ only in a few very minor respects from the actual world at t_0 – and there may well be amongst these worlds a world in which Austin succeeds in holing the putt. In this case, Dennett argues, we should say that Austin could have holed the putt. And indeed, Dennett goes on to suggest that not only need we not choose the narrow method – we positively ought to choose other methods when we are considering possibility claims – for it is only by looking not at identical but rather at closely similar worlds that we achieve any modal understanding at all. Otherwise we are condemned to the futility of actualism – constrained to point out, for instance, in the face of the claim that a steamship now travelling at 20 knots can do 25 knots that strictly speaking, with conditions as they in fact are, the steamship is capable only of the exact speed she is in fact attaining.¹¹

If this is true, then there will be a perfectly acceptable sense in which some of the fates avoided by the avoiders in Conway's Life World were fates which were, for those avoiders, *possible* fates (for they are fates to which these avoiders succumb in possible worlds which differ only in minor respects from the actual world in which they are avoided). In that case, something like the following version of Dennett's argument might look as though it were available:

P1 In some deterministic worlds, there are avoiders, avoiding harms.

C1 Therefore, in some deterministic worlds, some things are avoided.

P2 Whatever is avoided is avoidable, or evitable.

P3 Whatever is avoidable could have come about.

C1* Therefore, in some deterministic worlds, some avoided things (avoided fates) could have come about.

P4 In situations where avoided fates could have come about, not everything that happens is inevitable.

C2 Therefore, in some deterministic worlds, not everything is inevitable.

C3 Therefore, determinism does not imply inevitability.

For it would seem as though Dennett accepts that there is a sense of claim P3 on which it is true – the sense in which the answer to the question what could have happened in some situation is not determined by the narrow method. And so provided we are inclined also to think that P4 is plausible, it might look as though we are home and dry.

Before romping home to safety, though, it is worth stopping to reflect on how the definition of determinism which Dennett borrows from Van Inwagen is to be squared with all this. Determinism, recall, is said by Dennett to be the thesis that there is only one physically possible future. But how are we supposed to make that idea compatible with our new intermediate conclusion, C1*, that there are deterministic worlds in which other, non-actual fates could have come about? If at any instant only one future is physically possible, how can it also be true that other futures could also have unfolded, futures in which certain creatures go on to succumb to fates to which they do not succumb in the actual world? Dennett's answer, presumably, is going to have to be that contradiction is avoided by virtue of the fact that each possibility claim is relative to a different set of possible worlds. In saying that only one future is physically possible, we consider only possible worlds which are identical in their starting conditions to the actual world, and which share its laws; in saying that other futures could also possibly unfold, we consider also worlds which are only very similar to the actual one. Thus, in one sense, only a single future is possible under determinism; in another sense, many are available. It all depends on your choice of possible worlds. So far, so good. We could use a subscript – 'B' for a broad and 'N' for a narrow method of possible world selection - in order to indicate which sense of possibility was in question, so that the argument would now go as follows:

P1 In some deterministic worlds, there are avoiders, avoiding harms.

C1 Therefore, in some deterministic worlds, some things are avoided.

P2 Whatever is avoided is avoidable, or evitable.

P3 Whatever is avoidable is a fate which could_B have come about.

C1* Therefore, in some deterministic worlds, some avoided fates could_B have come about.

P4 In situations where avoided fates could_B have come about, not everything that happens is inevitable.

C2 Therefore, in some deterministic worlds, not everything is inevitable.

C3 Therefore, determinism does not imply inevitability.

But the question now is where this leaves us with respect to inevitability. Shouldn't we introduce a subscript for the concept of inevitability, too? For one might well think that the connections between the concepts of possibility and inevitability are such that the very same indeterminateness exists with respect to the concept of inevitability as has been conceded to arise in the case of possibility; indeed, does not the claim that a certain outcome was "inevitable" mean, roughly, that it was impossible that anything else should have happened? But in that case, the possible-worlds-related ambiguities which infect possibility claims will affect inevitability claims, too. Thus, instead of

arriving, as Dennett hopes to do, at an unqualified version of C3, we will be able to arrive only at the conclusion that one can give a *sense* to certain claims to the effect that such and such an actual outcome was inevitable, such that determinism does not imply that this is so – that determinism does not imply inevitability_B. But this, of course, leaves it looking obvious that there is another sense in which determinism *does* imply inevitability – the sense which is given by choosing the narrow method of selection for the set, X, of relevant possible worlds. In considering possible objections to his argument, Dennett has Conrad, his imaginary interlocutor, insisting forlornly that he still feels that there must be *some* sense of “inevitable” such that what happens in a determined world is inevitable. Dennett pretends that the relevant sense is elusive and that the burden of proof is on Conrad to produce it. But it seems to me that it is the easiest thing in the world to produce it, once one has conceded that determinism is to be defined as the claim that there is at any instant exactly one physically possible future. If there is exactly one physically possible future, then it would seem that all other futures are physically impossible – which is as much as to say that the way things in fact go is indeed inevitable on one perfectly intelligible understanding of that claim – the actual course of events is inevitable_N.

It must be, I think, that Dennett hopes that the etymology of “inevitable” will do some work to rule out this reading. That he does indeed suppose it can bear some weight is suggested by the following passage:

If “inevitable” is not just a synonym for “determined”, what does it additionally convey? Inevitable outcome? Inevitable by whom? Inevitable by the universe as a whole? That makes no sense, since the universe isn’t an agent with an interest in avoiding anything. Inevitable by anybody? But that is false; we’ve just seen how to distinguish the skill-ful avoiders from their less talented kin in some deterministic worlds.¹²

But Dennett does not explain why, exactly, we have to take the etymology of ‘inevitable’ seriously – any more seriously than we should take the etymology of any other word. The meanings of words evolve and change in response to the demands we make of them – why is it any more sensible to insist that we may not use ‘inevitable’ even where we cannot identify any particular agent in respect of whom some fate is unavoidable, than it is to insist that we may not use ‘impossible’, which, after all, derives from a verb meaning ‘to be able’, to characterise impersonal states of affairs as well as feats? Dennett’s answer will have to be, I assume, that there is nothing left to mean by “inevitable” if it does not mean “unavoidable by some particular agent or agents” – but he does not really explain what there is standing in the way of the very simple idea that there is an impersonal conception of inevitability which has evolved from the agent-based one and by means of which we can express such natural thoughts as these: that things were bound to go just as they did in the Life World, that the events which occurred there simply had to do so; that there were no other possibilities.

Perhaps Dennett thinks that determinism ceases to be worrying for us, provided we can be reassured that at least it does not imply inevitability_B. But it is just not clear that it is reassuring *enough* to be told that at least there is a *sense* in which things could have gone differently, as long as it remains clear that there is also a sense in which they could not (the sense which Dennett thinks is given by choosing the ‘narrow method’) – especially since what this seems to amount to is the claim that things could have gone differently *if* conditions had been ever so slightly altered, but had to go just as they did, given conditions as they precisely were. For one might reasonably think that what we are most concerned with when we fear that

determinism threatens our freedom is the question what we could and could not have done in the actual situation – not what we might have been able to do had things been somewhat changed. We are owed, at least, an account of why we need not worry that an explanation which admits that things could have gone differently only under changed conditions is *irrelevant* to our worries about the lack of freedom under the supposition of determinism – and Dennett does not provide one.

I have, then, some serious reservations about Dennett's argument. But in some ways, I am not at all out of sympathy with one of his main ideas – namely, the thought that the so-called 'narrow method' is not, on the whole, a useful method of selecting the possible worlds by reference to which we are to consider many everyday possibility claims. Dennett claims, indeed, that "the sense of 'can' invoked in these uncontroversial claims ... is one that *requires* us to look not at "conditions as they precisely were" but at minor variations on those conditions".¹³ I should not put it quite like that, because, as will shortly become clear, Dennett's account of how consideration of possible worlds enters into these everyday possibility claims seems to me flawed. Much of the next section of the paper will be devoted to explaining what is wrong with it. But once we have developed a (hopefully) improved explanation of how consideration of possible worlds might enter into such judgements, I think it will be possible to see a means of arguing for a view which is close in spirit to Dennett's claim about the requirement not to use the narrow method when assessing such judgements.

5. Possibility and Possible Worlds

The problem with Dennett's argument arises, in effect, because he is not even-handed in his treatment of the concepts of possibility and inevitability. He hopes to exploit the idea that the etymological connections between the concept of inevitability and the idea of avoidance might be used to insist, in effect, that 'inevitable' is always to be interpreted as 'inevitable_B' and thus to forge an objection to the suggestion that in deterministic worlds, whatever actually happens is inevitable. I have argued that it is implausible that etymological considerations can bear this kind of weight, and so that we must concede that there remains a perfectly good sense in which outcomes are all inevitable – perhaps we might say that they are inevitable_N - under determinism. But in fact, it seems to me that the possible worlds framework which Dennett seeks to employ in order to explain how an outcome might be considered impossible from one point of view, yet possible from another, is in any case misguided. Recall that his idea is that judgements such as Austin's claim that he could have holed a certain putt are assessed by means of a process that involves the selection of a set, X, of possible worlds which we are to 'canvass' in order to see whether the judgement is acceptable. But is it really true that any such canvassing of possible worlds always enters into these ordinary, everyday judgements about possibility? Suppose, for example, reflecting on some recent journey, and rueing a rash route-finding decision, I say "I could have turned left at that junction instead of right!" What sort of thinking is it that I have to engage in, in order to make the judgement (or in order to assess it when it is made by someone else)? Surely it is only this: I think back to the time shortly before I made the decision to turn right; I remember that there were no special obstacles to my turning left, no signs to indicate that this would be an illegal manoeuvre, no one with a gun to my head insisting that I turn right, etc. And I conclude straightforwardly that I could have turned left. Where do possible worlds come into this? It is surely the

actual world at a particular point in time that I consider – there is simply no ‘canvassing’ of possible worlds of the kind that Dennett imagines. And though the case raises additional complications, of which more in a moment, one might very well argue that the same is also true of Austin’s judgement that he could have holed that putt. Does Austin really need to ‘canvass’ any possible worlds in order to come to this conclusion? Is it not rather the case that he has to think only of the actual world – and of such things as his general skill as a golfer, his form on the day, and the circumstances he was facing at the time of the fateful putt (e.g. short putt, smooth green, light winds)?

Why, then, does Dennett suppose that there this canvassing must take place? The main reason, I think, is probably the fact that there are many examples – including the Austinian one on which Dennett focuses – which invite the thought that we cannot make a judgement of the ‘could have done otherwise’ kind without there being an implicit ‘if’ clause lurking somewhere in the background, which clarifies for us what kind of alteration in background conditions we are supposed to be envisaging. We often allege that there are things that we could have done under certain conditions, which we could not have done under certain others – certain others which might include the actual conditions. I could have picked up that 100kg bar if I’d trained for six months beforehand, perhaps, but not if you’d simply asked me right off to do it there and then. I could have opened the door if the bolt hadn’t been so rusty (I had the key, after all!), but not given the way it actually was. Austin could have holed the putt if that duck hadn’t distracted him by quacking at the crucial moment, but he couldn’t have holed it, given the off-putting noise. And so on. In these sorts of examples, it might seem as though what I have to do in order to make or assess a ‘could have done otherwise’ judgement is precisely what Dennett suggests I have to do – that is, select some set of worlds relative to which I am to consider the question. In worlds where no duck quacks, for instance, or where he remembers to relax his shoulders, or where he has practiced his putting the previous weekend, Austin could have holed the putt. With conditions as they actually were, however, we must concede that he could not have done.

Tempting though this account is, though, it cannot be right. One must not of course deny the phenomenon from which it draws its plausibility; it is indisputable that we often combine ‘could’ and ‘if’ in the ways evidenced by these examples. What is wrong is the suggestion that the possible worlds framework can be used in quite the simple way Dennett supposes to understand what is going on when we do this. Let me say at once that I am not implacably opposed to possible worlds; I am quite happy to concede, at any rate for the purposes of argument, that talk of possible worlds can be a useful heuristic device for thinking about certain sentences which are close relatives of those we are here considering – namely, sentences alleging that certain things *would* have happened, if certain conditions had been fulfilled (it is these sentences which are normally the target of treatments of so-called ‘counterfactuals’). In these cases, it is not at all unattractive to describe the imaginative activity we go in for in order to assess these claims as a ‘canvassing’ of possible worlds – typically, we canvass those which are close to the actual one but differ from it in some respect characterised by the if-clause, and we look to see how we imagine those worlds as unfolding. Thus, for example, I might say ‘I *would* have turned left, instead of right, if I hadn’t had the map upside down’, imagining how various possible worlds in which I have the map the right way up unfold. But ‘could ... if’ and ‘would ...if’ cannot be treated alike. With a ‘would ... if’ claim, we make what can seem to be a helpful exchange when, instead of wondering what *would have* happened in *this world* if

something *had been* otherwise, we consider what *does* happen in some *other* possible world where that thing *is* otherwise. The awkwardness of ‘would have’ seems thereby domesticated – the difficult tensed modal is banished entirely and its work done instead by some kind of quantification over close possible worlds. But even if we are content with this treatment of ‘would have’, no similar domestication can possibly work for ‘could have’.

Let us begin by considering an explicit ‘could ...if’ construction of the kind which most naturally invites thought of possible worlds. Take, for instance, the sentence:

- (1) Austin could have holed that putt if that duck hadn’t quacked.

Let us try thinking about this in terms of possible worlds. On Dennett’s view, it would seem, I am to assess this judgement by considering possible worlds close to this one in which the duck does not quack, and see what happens. Suppose, doing this, I imagine Austin holing the putt, instead of mis-hitting it, as he in fact did in the actual world. But this is the imaginative result that is supposed (on a broadly Lewisian framework)¹⁴ to underpin a judgement of

- (2) Austin *would* have holed that putt if the duck hadn’t quacked.

And to say that Austin could have holed the putt is not to say that he would have holed it. What, then, is the imaginative outcome which instead is to underpin (1)? Perhaps one in which there are some close worlds in which Austin holes the putt and some where he does not? But *that* imaginative outcome would seem rather to relate to something like

- (3) Austin might (perhaps) have holed that putt if the duck hadn’t quacked.

This may perhaps be *one* interpretation of ‘Austin could have holed that putt if that duck hadn’t quacked’ (to my ear, a stress on the *could* gives this interpretation) - but it is certainly not the most natural one. The most natural interpretation construes the sentence not as a past tense version of ‘Austin may (perhaps) hole this putt if that duck doesn’t quack’ but rather as a past tense version of ‘Austin *can* hole this putt if that duck doesn’t quack’. These are not the same judgement, as is clearly shown by the fact that the former could be acceptable, even if Austin has no skill in golf at all; the latter, though, seems to imply possession of an ability, a capacity, on Austin’s part. I will therefore call it the ‘ability interpretation’.

What about the suggestion that a canvassing which reveals many close worlds in which the duck does not quack and in which Austin holes the putt, and only a few of this variety in which he does not hole it, is what distinctively underpins the ability interpretation of (1)? This, unfortunately, remains inadequate. What this canvass result corresponds to is rather the judgement that

- (4) Austin would very likely have holed the putt, if the duck hadn’t quacked.

And one can imagine circumstances in which, although we would want to agree that Austin could have holed the putt, if the duck hadn’t quacked, we would not want to accept that he would have been likely to do so. (One could imagine, for instance, that it was the duck’s quacking that somehow triggered in Austin the desire to have a try at

the putt in the first place – in which case we shouldn't agree that he would have been likely to hole it if the duck hadn't quacked, but might be perfectly prepared to accept that he could have holed it.)

Of course, it may perhaps be that there is some canvass result I have not thought of which might yet underlie (1). But the onus is certainly on those who believe that we canvass possible worlds when we assess 'could have done otherwise' type judgements to explain properly which canvass results would justify which judgements. My claim is that there can be no swapping of 'could have ϕ -ed, if such and such had been the case' for 'did ϕ , in all/some (close) possible worlds where such and such was the case' – for this only generates (at most) 'would haves' and (perhaps) 'might haves', not the very distinctive 'could haves', crucial to discussions about freedom, in terms of which we speak about our abilities (e.g., 'Austin could have holed that putt') and opportunities (e.g., 'I could have turned left instead of right').¹⁵ And it is crucial to note that we do not advance one jot if we say that we canvass possible worlds in order to see what *could have* happened in those worlds. If this is the strategy, the awkward, unreduced modal simply remains as part of the proposition whose truth at the various possible worlds I am asked to 'canvass'. If I cannot tell whether Austin could have holed the putt in this world, without seeing whether he could have holed it in certain others, how on earth am I to tell whether he could have holed it in those others without doing further canvassing? Do we need to choose yet another set Y, of possible worlds to canvass, in order to determine whether he could have holed the putt in each member of the original set, X? Some 'could have' judgements, it would seem, have to be assertible without canvassing possible worlds, on pain of regress.

What, then, really *is* going on, then, when we say such things as (1) (intending the ability interpretation)? Clearly, the 'if'-clause¹⁶ does indeed demand that we do something which could be called a 'canvass' of those possible worlds which are much like the actual one but in which the duck does not quack – but what we must recognise, I think, is that the results of this canvass are not simply directly encoded by the 'could' statement, as Dennett's account suggests (that would deliver a 'would' or a 'might' judgement – not a 'could' judgement). Rather, the *main* business of the sentence is to attribute an ability to Austin, and the purpose of the canvass is to see whether the ability meets the following condition: would its possession by Austin on the relevant occasion have resulted in his holing of the putt if he had tried to do so, and if the duck had not quacked? What the sentence means, I should say, is roughly: 'Austin had (at the relevant time) an ability or capacity of a kind that would very likely have resulted in him holing the putt if he had tried to do so, and if the duck hadn't quacked'. We must consult possible worlds, it is true, to see whether this judgment ought to be made. But the statement is not *about* possible worlds, except indirectly. Its main job is to attribute an ability to Austin on a particular occasion in the past – and therefore to say something about the *actual* world.

It might very reasonably be asked at this point whether the attribution of an ability is not really itself a possible worlds judgement in disguise. Is not the attribution of an ability to a person really a way of saying that they would very likely have succeeded in doing something if they had tried? It may be arguable that we should think of abilities in this way – and the idea that some such reductive account is plausible is no doubt what account for the constant temptations of analyses of 'S could have done otherwise' along these lines. But I think it is unarguable that the grammar of English reveals that we do not in fact think of them (or at any rate speak of them) in this way. 'Could' and 'can' (as used in the sorts of contexts currently

under consideration) seem to me to be the markers of a basic metaphysics which accepts that powers are, as it were, denizens of the actual world, things whose possession may indeed have consequences for judgments about merely possible worlds, but not things whose possession merely *consists* in the truth or appropriateness of such possible worlds judgements. And the proof of this, as Austin himself saw, is that ‘can’ and ‘could’ are used so often in their ability sense in judgements that are quite unmodified by any ‘if’ clause.

To see this, let us return now from our examination of (1) and consider instead the simpler claim that Austin himself discussed:

(5) Austin could have holed that putt.

Dennett’s treatment in effect suggests that Austin is mistaken if he believes that any such simple claim can be made - for on Dennett’s view, a ‘could’ judgement of this sort is always made relative to a choice of possible worlds. Some choices may be more natural than others, granted, but a choice there always has to be. In effect, this is to say, I think, that on Dennett’s view, ‘coulds’ are what Austin called constitutionally iffy; what is really asserted is always a conditional of the ‘if such and such had been the case, A could have ϕ -ed’ variety. Dennett will thus allow that Austin could certainly have holed the putt had conditions been slightly different (to use my notation, he could_B have holed it), but insist that if we choose the narrow method of selecting possible worlds, we have to concede that he could not have done (he could_N not have done). My verdict, though, is quite different. On my view, ‘coulds’ are *not* constitutionally iffy. What is asserted by Austin when he claims that he could have holed the putt, is simply that he had an ability which made likely his successful holing of the putt, on the occasion of his attempt to putt it. He did indeed have such an ability (since he was a good golfer, on form, the putt was short, the winds were light, etc.) – and although he did not in fact hole the putt, it is nonetheless straightforwardly true that he had an ability which made it likely that he would do so. He could indeed therefore have holed it. And this conclusion is not relative to any choice of possible worlds. It is instead, just as it looks, a claim about the actual world.

Could Austin have holed the putt, though, Dennett might press, *given conditions as they precisely were*? This is the question to which Dennett wishes to return the answer ‘no’, assuming that the phrase ‘given conditions as they precisely were’ must trigger the operation of the so-called ‘narrow method’. According to Dennett, in order to answer this question, we are to canvass possible worlds which are *identical* to the actual world to see what happens in them – and there is only one of those, the actual world itself – and we know what happens in that. Austin misses the putt. Conclusion: he could not have holed it, given conditions as they precisely were. But this should give us pause for thought. First, should not this ‘canvassing’ of possible worlds rather produce the conclusion that Austin *would* not have holed the putt, given conditions as they precisely were? As explained above, judgements about what *does* happen in certain possible worlds are supposed, on the standard Lewisian account, to stand proxy for statements about what *would* have happened if ..., not for statements about what *could* have happened if ... Granted, the claim

(6) Austin would not have holed the putt given conditions precisely as they were.

is very bizarre. When we try to decide what would have happened if such and such had been the case, we are usually considering counterfactual circumstances, and it is

not at all clear what to make of the suggestion that Austin would not have holed the putt if things had been exactly the same as they in fact were. But we should not allow its dubious grammaticality to deter us from recognising that if anything, it is (6), and not

(7) Austin could not have holed the putt given conditions precisely as they were.

that a possible worlds canvass of the sort described above ought to deliver. Where, then, does Dennett get (7) from?

I suspect Dennett falls naturally into the assumption that (7) must be what his canvassing of worlds supports, because unlike the question whether Austin *would* have holed the putt if conditions had been precisely as they were, which does not really seem to be intelligible at all, the question whether he *could* have holed the putt, given those same conditions, has a reading, at least, on which it makes sense – namely, the reading on which it is simply equivalent to the question ‘Could Austin have holed the putt?’. Often, we use the past participle ‘given’, not to fix conditions for the purposes of some imminent inspection of possible worlds, but simply to highlight a piece of information that we consider relevant to the judgement we are being asked to make. For instance, I might ask: ‘Could Slesarenko have won the high jump competition, given that he had a thigh strain?’ This is not (on the interpretation I intend) a different question from the question whether Slesarenko could have won the high jump competition – it merely highlights the thigh strain as relevant information and asks whether (now that we know he had such a thigh strain) we should judge that Slesarenko could have won the competition. And similarly, the question whether Austin could have holed the putt, given conditions as they precisely were, can simply be treated as equivalent to the question whether he could have holed the putt, simpliciter. But if we give it this interpretation, the answer to the question is simply ‘yes’ (on the suppositions made above about Austin’s skill, form and the conditions on the day), as I have already argued.

Is another interpretation of the question available, though? – one which might deliver (7) as its answer? The thought that such an interpretation *is* possible assumes that the work done by the ‘given’ clause is much the same as that done by the ‘if’ clauses we find in combination so often with the ‘could’ of ability – that is, that it fixes certain conditions which we are to hold imaginatively constant for the purposes of some ‘canvass’ of possible worlds, and that the difference between ‘if’ and ‘given’ is merely the trivial one that the former, but not the latter is normally used to introduce a *counterfactual* circumstance. But as I argued above, the canvass in the ‘if’ cases is conducted in order to see whether the ability which it is the main function of the sentence to attribute would likely have resulted in the relevant outcome, under the imagined circumstances – whether, for example, to take the case I discussed earlier, Austin would have been likely to hole the putt if the duck hadn’t quacked. The sensible conduct of the thought experiment involved therefore demands in the first place that the relevant ‘would’ question is in order. But the relevant ‘would’ question in the case under consideration would seem to be this: would Austin’s ability have been likely to result in the successful holing of the putt, had conditions been precisely as they were? And this is a question which, like (6), to which it bears an obvious close relation, seems out of order. It has the ‘would have .. if .. had’ form which is the marker of a counterfactual question, but the semantic content of the subsidiary clause conflicts with this signalling of a counterfactual supposition. We can ask whether Austin did in fact hole the putt; we can ask whether he could have holed it; we can ask

whether he would have holed it if things had been slightly different; and we can ask whether he could have holed it if things had been slightly different. But we cannot ask whether he would have holed it (or would have been likely to hole it) if things had been exactly as they were. Things *were* precisely as they were, so we can make no sense of the instruction that the ‘if’ clause here delivers for us. There is no space for this question: it simply collapses into the factual question of whether Austin did indeed hole the putt. And so there is no space either for the question whether Austin could have holed the putt, given that conditions were precisely as they were, except on the interpretation, explained above, which delivers the answer ‘yes’.

What I would like to suggest, then, is this. It is an illusion spawned by a very complex intertwining of confusing grammatical phenomena to suppose that we have to accept that Austin could not have holed the putt, given conditions as they precisely were. We should simply say that though he did not hole the putt, he could have done; and on the only interpretation of it which makes any sense, we must accept also that he could have done, given conditions as they precisely were. I do not straightforwardly accept Dennett’s claim that that “the sense of ‘can’ invoked in these uncontroversial claims ... is one that *requires* us to look not at “conditions as they precisely were” but at minor variations on those conditions”, because, as I have explained, I do not think the ‘canvassing’ we go in for in assessing possibility judgements has quite the shape that Dennett thinks it does. But nevertheless, the *spirit* of his claim is correct. When we use ‘can’ in the uncontroversial claims Dennett is thinking of, we are either simply attributing an ability¹⁷ which we judge would have been likely, on the given occasion, to result in a certain outcome; or (when ‘if’ is involved) we are attributing an ability which would have been likely to result in that outcome if certain things had been slightly altered. Making and assessing these modal judgements simply never requires us – and indeed, better – it does not *permit* us to make a possible worlds judgment which demands the ‘canvassing’ only of the actual world.

6. Does determinism imply inevitability?

It seems to me, then, that there is a better way than Dennett’s of insisting upon the simple truth¹⁸ of statements such as (5), and of refusing to accept that there is any allowable interpretation of those such as (7) on which they come out true, under the envisaged circumstances. But where does this leave us? We have, in effect, now rejected the framework which I suggested might be utilised by Dennett in order to dispel the appearance that C1* is self-contradictory, given Dennett’s definition of determinism, for we have rejected the idea that ‘could’ statements are constitutionally iffy in the way this framework implies. In the absence of this resource, I suggest we should simply embrace the obvious conclusion that C1* really *is* self-contradictory; and hence accept that it is not the case that there are deterministic worlds in which avoided fates could have come about. Since C1* is derived from P1, P2 and P3, at least one of these premises will need to be rejected. It seems to me that the choice is between P1 and P3, and that the decision between the two is largely terminological – we may describe what goes on in the Life World as a kind of avoidance, if we will, but then we will have to concede that we do not mean to imply, when we describe a fate as avoidable, to suggest that it could have come about; or we can insist that fates that are avoidable have to be possible, and hence that what goes on in the Life World is not genuine avoidance. But there is no room, without the help provided by the defective possible worlds framework, for Dennett’s attempt to have it both ways.

On the other hand, I have attempted to vindicate Dennett's suggestion that there is nothing wrong with our ordinary possibility judgements, and, in particular, that there can be no challenge to them from any application of the so-called 'narrow method'. The Life World may not contain alternative possibilities, but our world certainly does – that things could have gone differently from the way they in fact went and that the world might unfold from the present in numerous different ways are commitments which it seems to me cannot sensibly be given up without a wholesale attack on the very concept of possibility. Defined as the claim that there is, at any instant, exactly one physically possible future, determinism would seem to be rather straightforwardly inconsistent with the many everyday claims we make about what could and might have happened, what can be done and what may occur. We need a very good reason, therefore, for thinking that there is any chance it might be true. And I hope to have done something in this paper to undermine at least one kind of reason for supposing it might be – the suggestion that no one could ever have done anything other than he or she in fact did, *given conditions as they precisely were*. For this claim is either false (if it simply means that no one could ever have done anything other than he or she in fact did), or senseless (if *given* is treated as though it is merely a special kind of non-counterfactual introducing variety of *if*).

¹ London: Penguin, 2003.

² For the first, see his *Elbow Room: The Varieties of Free Will Worth Wanting* (Cambridge MA: MIT Press and OUP, 1984).

³ Oxford: Clarendon Press, 1983, p.3.

⁴ *Freedom Evolves*, p.25. Determinism can of course be defined in other ways; and not everything I say about it here may hold also for more traditional formulations.

⁵ p.58.

⁶ p.59.

⁷ As I use the term, a fate is to be specified by a 'that' clause – what is avoided is thus for example 'that Glider G is eaten by Eater E' or 'that Loaf L collides with Glider G'.

⁸ p.56.

⁹ This idea is developed a little further in Christopher Taylor and Daniel Dennett, 'Who's Afraid of Determinism? Rethinking Causes and Possibilities' in R. Kane (ed.) *The Oxford Handbook of Free Will* (Oxford: OUP, 2002): 257-77.

¹⁰ Austin 1961, p.166.

¹¹ p.77.

¹² p.57

¹³ p.77.

¹⁴ See his *Counterfactuals* (Oxford:Blackwell, 1973).

¹⁵ I say lamentably little about the latter kind of judgement in this paper, having found it necessary to focus, for want of space, on 'could' statements which attribute abilities.

¹⁶ I attempt to avoid talk of 'antecedents' and 'consequents' having been convinced by the work of V.H. Dudman that the conditionals in question contain no such things. See his 'Antecedents and Consequents', *Theoria*, 52 (1986): 168-99.

¹⁷ Of course, there are 'can' statements which attribute opportunities rather than abilities – and indeed there are other kinds of 'can' statement, as well. But there is not space here to give separate accounts of all these different varieties of judgement.

¹⁸ In the envisaged circumstances - i.e. short putt, light winds, Austin a good golfer on good form, etc.

References

Austin, J.L. (1961) 'Ifs and Cans' in his *Philosophical Papers* (OUP: Oxford): 205-32.

-
- Dennett, D. (1984) *Elbow Room: the Varieties of Free Will Worth Wanting* (OUP and MIT Press: Oxford and Cambridge, MA).
- Dennett, D. (2003) *Freedom Evolves* (Penguin: London).
- Dudman, V.H. (1986) 'Antecedents and Consequents', *Theoria* 52 (1986): 168-99.
- Lewis, D. (1973) *Counterfactuals* (Oxford: Blackwell).
- Taylor, C. and Dennett, D. (2001) 'Who's Afraid of Determinism? Rethinking Causes and Possibilities' in R. Kane (ed.) *The Oxford Handbook of Free Will* (OUP: Oxford and New York), 257-77.
- Van Inwagen, P. (1983) *An Essay on Free Will* (OUP: Oxford).