The Problem of Fr** W*ll

But who sees not that all the dispute is about a word?
                                            Berkeley

A couple of years ago, I was invited to give a talk on the topic "A Philo-sophical Perspective on Free Will" at a "summer school" on science and religion – an occasion at which I was the only philosopher who spoke, the other speakers being scientists and theologians. I didn't choose the title of my talk. It was, rather, proposed by the organizers of the summer school, and I reluctantly accepted it. I concede that it wasn't a bad title for practical purposes; that is, it indicated in a very loose sort of way what I intended to talk about. But I disliked it. In fact, I disliked it intensely. I disliked it because it implied something that I think is false, namely that there's some reasonably well-defined thing called 'free will' and that specialists in various studies or sciences or disciplines have, or might be expected to have, different "perspectives" on it: physicists view (or might be expected to view) this thing called 'free will' from one perspective, neurobiologists from another, philosophers from a third perspective, theologians from a fourth – and so on.

Let me consider a contrastive analogy – an imaginary interdisciplinary conference on some topic on which the representatives of different disci-plines might be expected to have distinct "perspectives." Let us imagine a gathering of representatives of different fields of study – let us say, astron-omy, biology, communications engineering, and theology – to discuss the topic "extraterrestrial intelligent life." The idea in the minds of the organ-izers of the conference is that the astronomers present will discuss extrater-restrial intelligent life from one perspective, the biologists from another, the communications engineers from a third, and the theologians from a fourth. The idea of such a gathering is not at all puzzling, for – even if there are disagreements about the precise definition of 'life' or about what, exactly, 'intelligent' means – we have at least a rough-and-ready understanding of

the phrase 'intelligent life', and we know what it would be to bring the perspective provided by a particular discipline to a discussion of its existence elsewhere in the universe. (We would, for example, expect the astronomers at our imaginary gathering to talk about such matters as the proportion of stars that have planets on which life is possible and the kind of observations that might indicate that a planet was inhabited by an industrial civilization, and the communications engineers to talk about the problems of sending and receiving signals across intersidereal distances and about what features would mark a signal sent by intelligent beings.)

In my view, however, there is no analogy between this wholly unpuzzling case (*sc*. of the representatives of various disciplines coming together to discuss some thing or phenomenon from different perspectives) and the idea that representatives of various disciplines might come together to discuss "free will," each from the perspective of his or her own discipline. There is no analogy between the cases because the phrase 'free will' has no agreed-upon meaning. And I don't mean that there are rival definitions of 'free will' in the way that there are rival definitions of 'life' or 'intelligence'. The case of 'free will' is much worse than that. Everyone agrees that dogs and spiders and paramecia and algae are living things, and that rocks and steam engines and neutron stars aren't. Everyone agrees that bus drivers and mathematicians and lawyers, however stupid some among them may be, exhibit "intelligence" in the sense the word has in the phrase 'The Search for Extraterrestrial Intelligence'. And everyone agrees that mice and giraffes don't. That is to say, there are perfectly clear cases of things that are "alive" and "intelligent" and there are perfectly clear cases of things that do not enjoy these distinctions, and disputes about the exact meanings of these two words are just that: disputes about the *exact* meanings they do – or should – have. But no one has any idea, any idea at all, what 'free will' means.

We may distinguish two classes of people who have no idea what 'free will' means (I mean, two classes of people among those who talk about "free will" – in the sense of engaging in discussions in which the phrase 'free will' plays some important role). I'll call these two classes of people "the scientists" and "the philosophers." (Of course, most scientists – and, for that matter, most philosophers, at least in their published work – manage to get through their careers with out mentioning free will at all.) While neither group has any idea what 'free will' means, this is true of the two groups in different ways. Scientists who discuss "free will" use the term without any attempt at a definition – or else they provide some useless piece of verbal hand-waving that they seem to regard as a definition. (In a couple of cases, they provide definitions that are clear enough but simply

bizarre.) And that's also true of many philosophers. The people I'm calling "the philosophers," however, do provide reasonably precise and intelligible definitions of 'free will'; the trouble is, they don't all provide the same one. In fact, they provide wildly different ones. And then, to their shame, they go on to *argue* about who has the *right* one – or, as they tend to put it (again to their shame), they argue about "what free will is." And since the words 'free will' don't mean anything in particular, a dispute about "what free will is" is simply absurd. By way of analogy – if you can abide another analogy, so close on the heels of the last – one might imagine a rancorous dispute between some physicists (on the one hand) and some diplomatic historians (on the other) about what "force" is; the historians say that force is the application by a nation of military action to resolve a conflict that cannot be resolved by diplomatic means, and the physicists say that force is a vector: the net force acting on an object at an instant is the rate at which its momentum is changing at that instant. If you can imagine such a dispute, try your imagination further: imagine someone's describing it by saying that it shows that diplomatic historians and physicists "have different concepts of force."

　　Now you may want to tell me that what I'm saying is obviously wrong, since we do use the words 'free will' in everyday life, and we understand them as well as we understand any words we use. I reply that there is a sense in which it is true that we use the words 'free will' in everyday life (and use them in a perfectly intelligible way), but that this fact does not contradict what I have been saying. In everyday life – in the business of getting and spending, falling in love and raising families, voting in elections, consulting doctors and lawyers, and so on – when we use the words 'free will' it is almost always inside the phrase 'of one's own free will' (where 'one's' represents the position of a possessive pronoun). And there's no real dispute about what that means: if, for example, you're asked in a court of law whether you did something or other of your own free will, you're being asked whether you acted under duress – under any sort of coercion. (Were you, for example, threatened with certain untoward consequences if you did not act that way?) And everyone knows that people sometimes do things when they believe that their not doing them would have no untoward consequences. It is obvious that none of the disputes about "free will" that are so notable a part of the intellectual landscape, past and present, are about whether people ever do things without being coerced. Suppose someone asks,

How can we possibly have free will if God foresees everything we do?

Or, again, suppose someone asks,

How can we possibly have free will if a being with infinite power of calcula-tion and a knowledge of the laws of mechanics and of the forces that parti-cles exert on one another and a knowledge of the position and momentum of every particle of matter at any given time could calculate the position and momentum of every particle of matter at any other time?

Or, finally, suppose someone asks,

How can we possibly have free will if the Libet experiments show that physical conditions sufficient for our so-called voluntary bodily move-ments exist prior to our conscious decision to make those movements?

It is obvious that none of these three speakers is asking how it could be that our acts are uncoerced in the circumstances they have specified – for those circumstances don't even *seem* to be incompatible with the absence of coercion.

The ordinary meaning of the phrase 'free will' – or '*freier Wille*' or '*libre arbitre*' or what have you – simply does not explain its use in contexts like these. (Incidentally, everything I say about 'free will' will apply to any closely related words or phrases – such as 'freedom', 'free action', 'free choice' …)

"But what about your own work, van Inwagen?" the Interlocutor asks. "After all, you've written a book called *An Essay on Free Will*, and the words 'free will' occur frequently in many of your essays." Point well taken. And here is my answer to this well-taken point. In the seventies and early eight-ies, when I was doing most of my work on, well, free will, I was a naive product of my philosophical education. I was working within a philosoph-ical tradition or paradigm or whatever you want to call it that I had been initiated into by my teachers Richard Taylor and Keith Lehrer – who had been in their turn initiated into it by *their* teacher, Roderick Chisholm – in which the words 'free will' had a very specific technical meaning. I sim-ply supposed that the meaning 'free will' had in this tradition was the meaning with which all analytical philosophers used this phrase. And, oddly enough, this was, in those days, very close to the truth. Consider, for example, David Lewis's characteristically splendid paper "Are We Free to Break the Laws?"[1] – it was published in 1981 – a profound analysis and

---

[1] David Lewis, "Are We Free to Break the Laws?," *Philosophical Papers, Volume II* (Oxford University Press, 1987), pp. 291–298. The paper first appeared in *Theoria* 47 (1981): 113–121. Citations are from *Philosophical Papers II*. It is available online at www.andrewmbailey.com/dkl/Free_to_Break_the_Laws.pdf

criticism of the argument of my paper "The Incompatibility of Free Will and Determinism."² In his discussion of my argument, Lewis – without question or reservation or apology – used the words 'freedom' and 'free' (said of both an act and an agent) and the phrases 'free to', 'freely does', and 'act freely' in precisely the senses in which I had used them. But if it was ever the case that all or almost all analytical philosophers used these words and phrases (and the phrase 'free will') in the senses that I had supposed were their universally accepted senses, it is no longer. And it had not always been the case. If I were writing the book now, I would not identify its subject matter by reference to "free will" or by the use of any other phrase containing 'free' or 'freely' or 'freedom'.

I now turn to a statement of what seems to me to be a philosophical problem of great significance. This problem is intimately related to what *some* writers have meant by "the problem of free will." It is, in fact, closely modeled on a problem that I myself once proposed as a referent for 'the problem of free will'.³ You will notice, however, that the adjective 'free' does not occur in my statement of the problem – nor does the noun 'freedom' or the adverb 'freely'. My statement of the problem will be, so to speak, 'free'-free. If I sometimes *mention* the f-word, generally in the course of quoting the writings or discussing the work of others, I shall at any rate not *use* it again. (And I will not use the noun 'will' again, either.) You will also notice that there will be no mention of the moral responsibility of agents for their actions in my statement of the problem. I shall indeed speak of certain states of affairs *being the fault of* various agents, or, alternatively, of those agents *being to blame for* those states of affairs, and fault and blame certainly bear some intimate relation to moral responsibility, whatever the words 'moral responsibility' may mean. I take it that a statement like

> Jerry is (morally) responsible for his wife's unhappiness

can be nothing more than a philosopher's needlessly technical way of saying 'Jerry is to blame for his wife's unhappiness' or 'It's Jerry's fault that his wife is unhappy'. (I'm not much interested in what a philosopher might mean by saying that someone – Jerry, say – was "morally responsible" for some *good* state of affairs. I suppose it would be something like "Jerry gets the credit for it.") So I propose to speak of fault and blame instead of moral responsibility. And note that if one has decided to speak of fault and blame (as opposed to moral responsibility), and if that decision leads one to reflect on how judgments of fault and blame are framed, one will

---

² *Philosophical Studies* 27 (1975): 185–199.   ³ See Chapter 10 of the present volume.

realize that the objects of those judgments are states of affairs and not actions. That is to say, the things that can be said to be someone's fault (or for which someone can be blamed) are states of affairs that are in some way or other causally related to that person's actions (or inactions), not the actions themselves. The statement "It's Fred's fault she was elected; he cast the deciding vote" makes perfect sense. But what could a statement like "It's Fred's fault that he raised his hand when the chair said, 'All in favor?'" mean? In the unlikely event that someone did say that, I'd have to cast about for an interpretation; I suppose I'd decide that the speaker must have regarded some consequence of the hand-raising as a bad thing and had chosen a rather puzzling way of saying that that bad thing was Fred's fault.

Finally, I will not use the treacherous – because radically ambiguous – phrase 'could have done otherwise'.[4] (I will, in fact, not use 'could have' at all.)

So much for the words and phrases that will not occur in my statement of the philosophical problem I have promised you. Now the problem.

Four Theses

I begin by stating four theses. I am not affirming these theses; I am rather stating them so that I can refer to them in the statement of the problem. Two of the theses have familiar names: 'determinism' and 'indeterminism'. I'll state the other two first, however. These theses have no "standard" names, so I shall have to invent my own names for them. In order to ensure that these names are not tendentious, I will call them simply 'Thesis One' and 'Thesis Two'. Here is Thesis One:

> On at least some occasions when a human agent is trying to decide between two or more incompatible courses of action, that agent is able to perform each of them.[5]

---

[4] See Chapter 4 of the present volume.

[5] Alfred R. Mele has suggested to me that I should say something about the ambiguity of 'able to'. And this was a useful suggestion, for the phrase has many senses. In the text, I alluded to the "radical ambiguity" of 'could have done otherwise', and it is my firm opinion that 'was able to …' is less *dangerously* ambiguous – less likely to slip from one of its senses to another in the course of a philosophical argument – than 'could have …' Nevertheless, the phrase 'is able to' (whatever its tense; whatever the infinitive it governs) *is* ambiguous. For example: Grisha Sokolov has been stranded on a desert island; is he able to play the piano? In one sense, Yes, in another, No. Or: The loan officer at the First National Bank knows that she would lose her job if she approved your application for a loan; is she able to approve it? Of course: she has only to sign this piece of paper; and yet she tells you, "I'm afraid I'm unable to approve the loan you've applied for." Is she mistaken? Lying? I have discussed ambiguities of these and various other kinds that attend the phrase 'is able to' in section 1.4 (pp. 8–13) of *An Essay on Free Will* (Oxford: Clarendon Press, 1983). But the sense

If, for example, the following story is true, there has been one such "occasion." Early last January, Sally was admitted both to Julliard to study piano and to the Harvard Law School. At that time she wanted very much to become a concert pianist (for reasons of personal fulfillment). She also wanted very much to become a lawyer (in this case, her reasons were moral and political). She spent the month of January trying to decide whether to study piano at Julliard or law at Harvard (or, more immediately, whether to accept the Julliard offer and decline the Harvard offer or to accept the Harvard offer and decline the Julliard offer). At every moment during the course of these deliberations, she was able to do this:

> accept the Julliard offer and decline the Harvard offer,

and she was able to do this:

> accept the Harvard offer and decline the Julliard offer.

That is to say, at every moment in the course of her deliberations she had *both* those abilities. (Of course, having both the ability to do A and the ability to do B is not the same thing as having the ability to do both A and B.)

And here is Thesis Two (note that Thesis Two refers to Thesis One):

> If the bad consequences of a decision are ever the fault of the person who made the decision, then Thesis One is true.

An alternative formulation of Thesis Two:

> If anyone can rightly be blamed for the bad consequences of some decision he or she has made, then Thesis One is true.

---

of 'is able to' that figures in the argument of this paper may be specified by a simple device – by considering what is involved in being in a position to make a promise. Suppose that Alice asks Tim to give her a ride to work the next day (it's a serious matter; she'll lose her job if she counts on Tim for a ride and he fails to provide it). A necessary (and I think sufficient) condition for Tim's being in a position to promise to give Alice the requested ride is that he believe that he *is able to* give her a ride. And those italicized words have, in that context, the sense I mean 'is able to' to have in the argument in the text. Suppose Winifred and Sokolov are both castaways on the same island; able though he is to play the piano (in one sense of 'able'), he is not in a position to promise Winifred that he will play the piano that evening. And the loan officer is no doubt in a position to promise you to approve the loan ("no doubt": it might be that she is unsure whether it is psychologically possible for her to sign the piece of paper in those circumstances) – although of course it would be either foolish or dishonest of her to *make* such a promise. Having said all these things in response to Professor Mele's suggestion, I must concede that he will probably not regard what I have said as satisfactory. See his closely reasoned paper "Agents' Abilities," *Noûs* 37 (2003): 447–470. There are many points in this paper in which he and I are in fundamental disagreement, and an adequate discussion of them would require a separate paper. For a more extensive discussion of these points, see Chapter 14 of the present volume.

Consider, for example, the following story of a decision that has had bad consequences. One of Frank's students offered to have sexual intercourse with him if he would give her an A in his ethics course. Frank thought it over and decided to accept her offer – a decision that led to his losing his job, his family's being in serious financial need, and his wife's being driven nearly mad with rage and jealousy.

Now most people, on hearing this story, would say that all these bad things – Frank's losing his position, his family's severe need, his wife's near madness – were *his fault*. They would say that he was to *blame* for them. Let us suppose that most people are right: these things were Frank's fault and he can rightly be blamed for them.

Thesis Two implies that it follows from these bad consequences of Frank's decision being his fault that Thesis One is true. It follows, that is, that it is false that it is *never* the case that when a human agent is trying to decide between two or more incompatible courses of action, that agent is able to perform each of them.

We next state the other two theses, determinism and indeterminism, that will figure in the statement of the problem.

> *Determinism* is the thesis that the past and the laws of nature determine a unique future.

> *Indeterminism* is the thesis that the past and the laws of nature do not determine a unique future.

We now proceed to a description of the dialectical situation that, as it were, generates the problem.

The Dialectical Situation

There are seemingly unanswerable arguments that (if they are indeed unanswerable) demonstrate that Thesis One is incompatible with determinism. I allude, of course, to the various versions of the Consequence Argument, as it is known in the trade. And there are seemingly unanswerable arguments that (if indeed …) demonstrate that Thesis One is incompatible with indeterminism (this part needs a little work, since indeterminism does not imply that a *given person's* actions are undetermined; the work can be done). I allude, of course, to the various versions of the *Mind* argument – named in honor of the august journal in which so many variants on it have appeared. But if Thesis One is incompatible with both determinism and indeterminism, then Thesis One is false – necessarily false, in fact.

There are, moreover, seemingly unanswerable arguments that, if they are correct, demonstrate the truth of Thesis Two. (But what about Frankfurt's refutation of the Principle of Alternative Possibilities? Has Frankfurt not shown that – or at any rate, can his arguments not easily be adapted to show that – a certain state of affairs can be someone's fault even if no one trying to decide whether to do A or to do B has ever been able to do A *and* able to do B? I can only say that I explained many years ago why Frankfurt's arguments do not show any such thing, and that if people have not been listening, it's *not my fault*.[6] But this is really beside the point. I am stating a problem, not discussing possible solutions to that problem. And I am not stating this problem with an eye towards presenting a solution of my own to it. Frankfurt's arguments are best looked at as a proposed solution to the problem I am in the process of stating. Similar remarks apply to supposed refutations of the Consequence Argument and the *Mind* Argument.)[7]

But if Thesis One is false and Thesis Two is true, then nothing is ever anyone's fault. And it is evident that it is simply *false* that nothing is ever anyone's fault. It must, therefore, be that at least one of the following four propositions is true:

> The seemingly unanswerable arguments for the incompatibility of Thesis One and determinism are in fact answerable; these arguments are fallacious

> The seemingly unanswerable arguments for the incompatibility of Thesis One and indeterminism are in fact answerable; these arguments are fallacious

> The seemingly unanswerable arguments for Thesis Two are in fact answerable; these arguments are fallacious

> It is *not* evident that it is simply false that nothing is ever anyone's fault; and not only is it not evident, it's not even *true*: the apparent self-evidence of that thesis is illusory.

The Statement of the Problem

My statement of the problem is in the form of three interrelated questions:

---

[6] See Section 5.3 of *An Essay on Free Will* (Oxford: Clarendon Press, 1983) and Chapters 1 and 6 of the present volume.

[7] This may be a difficult exercise, but do your best not to attend to the question whether the statements I make in the course of laying out this problem are true or false; what you should be attending to is the fact that, in making these statements, I never use certain words and phrases, to wit: 'free', 'freely', 'freedom', 'could have done otherwise', and 'moral responsibility'.

> Which of these four propositions is true? If any of the first three
> is true, what are the fallacies in the arguments to which those
> propositions allude? If the fourth proposition is true, what is the
> nature of the illusion that has made it seem self-evident to me and
> many other philosophers (and, indeed, to the great mass of humanity)
> that many things that have happened in the course of human history
> are someone's fault?

For reasons that I hope I have made clear, I decline to call this problem 'the problem of free will'. It will, however, be convenient to have a name for it. I will call it the Culpability Problem – with the understanding that this name is a mere tag whose purpose is to facilitate reference – a proper name, if you will. The fact that *culpa* is the Latin word for 'fault' or 'blame' should be regarded as a mere *aide-mémoire*. (I introduce a name for the problem only at this late point in the chapter with the specific intention of underscoring the fact that I ascribe no significance to the name I have chosen. And if anyone does find some reason to dislike this name, if anyone regards it as tendentious or in any other way objectionable, I'll simply call it something else – 'the Three Questions Problem' or 'Peter's Problem' or 'Arthur'.)[8] Note that the only philosophical technical term that occurs in the statement of the Culpability Problem is 'determinism' – 'indeterminism' being merely the contradictory of 'determinism'.

I have given up on the Culpability Problem. It's too hard for me. But my purpose in this paper is not to solve it or even to examine proposed solutions to it. It's to show that much philosophical work whose announced subject is "the problem of free will" is simply irrelevant to the Culpability Problem – not addressed to that problem at all. (I don't deny that this work may be valuable for other reasons. After all, it's no objection to the discipline of social psychology that its investigations and theories are of no relevance to the problems of astrophysics.)

I do not mean to imply that *all* work that is specifically addressed to "the problem of free will" is irrelevant to the Culpability Problem. My own work is a case in point. The subject matter of, e.g., "The Incompatibility of Free Will and Determinism" and *An Essay on Free Will* is the Culpability Problem, although not under that name. The work on "free will" that is irrelevant to the Culpability Problem is that which involves what I shall call *verbal essentialism*.[9]

---

[8] Reporter: "What do you call that haircut?"; George Harrison: "Arthur."

[9] Well, much of it certainly does. Perhaps there is work on "free will" that is irrelevant to the Culpability Problem that does not involve verbal essentialism. I'll say this, at any rate: If some philosophical book or essay that purports to address "the problem of free will" does involve verbal essentialism (in the manner described below in the text), that is *sufficient* for its irrelevance to the Culpability Problem.

A piece of philosophical writing exhibits verbal essentialism if there is some philosophical term of art (either a word or phrase invented by philosophers – like 'actualism' and 'nowness' – or a dictionary-entry word that is used by philosophers in a special technical sense, like 'proposition' and 'validity') such that the thesis presented in that text could not be stated *without using that word or phrase*. I have, for example, recently accused Karen Bennett of falling prey to verbal essentialism – you will have guessed that I regard verbal essentialism as a Bad Thing, something one can properly be said to fall prey to – in her paper "Proxy 'Actualism'."[10] In that paper, Bennett presented certain criticisms of Alvin Plantinga's so-called actualism (in his philosophy of modality). In the introductory paragraph of my critique of Bennett's criticisms of Plantinga, I wrote,

> My conclusion will be that [Bennett's] criticisms fail, owing to the fact that they depend on the historical accident that the customary designation for Plantinga's position is "actualism" – that if this position had been given a name that did not contain 'actual' or any word formed from 'actual', the criticisms of the position that are presented in "Proxy 'Actualism'" could not even be stated.[11]

The works I mean to call attention to are like that, *mutatis mutandis*: you could not rewrite them in such a way as to eliminate the phrase 'free will' and the words 'freedom' and 'freely' from them – there would simply be nothing left; a translation of these works into 'free'-free language is impossible. (That would not be the case with, for example, *An Essay on Free Will* – which is essentially an essay on the Culpability Problem; nothing in its substantive content hangs on my choice of 'free will' as a term of art.)

It is time to turn to examples. I begin with Dan Dennett's latest thoughts on free will, those contained in the chapter on that topic in his recent book *Intuition Pumps and Other Tools for Thinking*.[12] But lest you think that my accusation of verbal essentialism is directed only at those philosophers who, like Dennett, take what might be called a "deflationary" position on "free will," let me assure you that I will also consider examples of philosophers (conveniently cited by Dennett) who speak favorably of things like "ultimate moral responsibility" and "absolute free will" and agents who are

---

[10] *Philosophical Studies* (2006): 263–294.
[11] "'Who Sees Not that All the Dispute is About a Word?': Some Thoughts on Bennett's 'Proxy "Actualism"'," *Hungarian Philosophical Review* 3 (2012): 69–81.
[12] New York: W. W. Norton, 2013.

"perfectly free to do otherwise." Those philosophers and Dennett, in my view, occupy two sides of the same coin – the same verbally essentialist coin.

One of Dennett's targets in that chapter is those scientists who contend that science has shown that free will is an illusion. He says of these scientists that they

> have typically been making a rookie mistake: confusing the manifest image [Dennett regards free will, like color and solidity, as a denizen of Sellars's "manifest image"] with what we might call the folk ideology of the manifest image. The folk ideology of color is, let's face it, bonkers; color just isn't what most people think it is, but that doesn't mean that the manifest world doesn't really have any colors; it means that colors – real colors – are quite different from what most folks think they are … Similarly, free will isn't what some of the folk ideology of the manifest image proclaims it to be, a sort of magical isolation from causation.[13] … I wholeheartedly agree with the scientific chorus that that sort of free will is an illusion, but that doesn't mean that free will is an illusion in any morally important sense. It's as real as colors …

and, moreover, that they

> are making the mistake people make when they say that nothing is ever [colored],[14] not really. They are using an unreconstructed popular concept of free will, when they should be adjusting it first, the way they do with color …

Now I am not convinced that science has shown us that color just isn't what most folks think it is. This is not because I think that what science has revealed about color shows us that color *is* what most folks think it is. It's rather that I'm not sure whether "most folks" think color is *anything*, bonkers or not – whether they have any particular views on what color is. I'm not sure whether there is a "folk ideology" of color. If there is such an ideology – I would ask – what is its propositional content? Dennett mentions the indisputable facts that electromagnetic radiation in the 390–700 nm range is not made of little colored things and that atoms have no colors, but I can't tell from the context of that remark whether he meant it to imply that the denials of these two propositions

---

[13] A very misleading phrase. See note 17.

[14] Dennett actually has 'solid' and not 'colored' here. I have substituted 'colored' for 'solid' in order that my scattered quotation should express a unified thought. Dennett had earlier used solidity and color as parallel examples of things that are *real* and yet not much like what we thought they were before science revealed their true nature.

were contained in the folk ideology of color. If he did, he would certainly be wrong, since this folk ideology of color, if it ever existed, must have antedated our knowledge of atoms in the modern chemical sense of the word and of electromagnetic radiation. (I don't deny that ordinary people often have *false general beliefs* about colors. My father, an art school graduate and a fine painter who knew all about how to mix pigments to obtain a desired color, firmly believed that green was a mixture of blue and yellow. And that's a general belief about color – at any rate, a general belief about three particular colors – that is just *not true*. There is a perfectly good sense in which orange is a mixture of red and yellow and purple is a mixture of red and blue, but green is not a mixture of blue and yellow in that sense or any other. My father's false belief couldn't be said to be bonkers, for it was based on a very extensive range of experience: there had been hundreds of occasions on which he had mixed a blue pigment and a yellow pigment, and every single time the resulting mixture was a green pigment. But that experience, extensive though it was, was deceptive: it was due to certain accidental physical properties of commercially available blue and yellow pigments. In my view, this wasn't a case of my father's having a false *theory* about color; he simply had a general belief about three colors – blue, yellow, and green – that was wrong.)

But let's suppose that Dennett is right: there *is* a folk ideology of color and scientific investigation shows that it's (almost?) entirely wrong; but we should not infer from the (almost) entire wrongness of the folk ideology that color is an illusion; color is real, but it's not what everyone used to think it was and most people still think it is.

But what, then, does it mean to say that color is real and why does Dennett think it is real in that sense? Dennett does not define 'real' and he does not exactly argue for the reality (in any sense) of color; what he says in defense of the reality of color is more along the lines of an appeal to common sense: "[Color] is not an illusion in the sense that matters: nobody thinks Sony is lying when it says that its color televisions really show the world of color or that Sherwin-Williams should be sued for fraud for selling us many different colors in the form of paint." I would myself offer something a little more theoretical in support of the reality of color – although what I would offer is certainly consistent with the point Dennett means those two examples to illustrate. I would say something along the following lines. Most of the statements we make in ordinary life that contain color-words ('The car that left the scene of the accident was a dark green Lexus'; 'Titanium dioxide is the most common white pigment'; 'The

predominant color of Picasso's *La Vie* (1903, Cleveland Museum of Art) is blue') are *true*;[15] and if the sentences 'Color is an illusion' or 'Color is not real' mean anything at all (they certainly don't wear their meanings on their sleeves), they must mean something that implies that most of those statements are false.[16] But that's by the way.

Dennett's thesis, then, is that free will is like color (as he represents color). It's real enough, but it can't be what the folk ideology of free will says it is: there's no place for *that* in the scientific image (the image that stands in opposition to the manifest image). The "unreconstructed popular concept of free will" is inconsistent with what science has discovered about the nature of the beings (us) to whom it is supposed to apply.

I said a moment ago (in effect) that I was not sure whether there was any such thing as the *unreconstructed popular* concept of color – although I was willing to grant for the sake of argument that there was such a thing. However that may be, there is certainly such a thing as the concept of color. That is to say, there are such things as the meaning of the word 'colored' and the meanings of the words 'green' and 'brown' and 'mauve'. (Let nominalists understand that statement as they will: there has to be some sense in which it's true.) Here is why the case of the concept of color (even assuming that Dennett is right when he says that there is an unreconstructed popular concept of color and that nothing in the real world corresponds to it) is not parallel to the case of the concept of free will: there is no such thing as the concept of free will. And, of course, if there is no such thing as the concept of free will, there is no such thing as the unreconstructed popular concept of free will.

But I suppose I'm getting ahead of myself. Before I say anything more about the non-existence of the concept of free will, I should tell you what Dennett thinks the unreconstructed popular concept of free will is, and I should tell you what the scientific-image-friendly concept with which he means to replace it is. As to the former, he says (towards the end of the chapter):

> People care deeply about having free will, but they also seem to have misguided ideas about what free will is or could be (like their misguided ideas about color …). Our decisions are not little miracles in the brain that violate the physics and chemistry that account for the rest of our bodies' processes, even if many folk think this must be what happens if our decisions

---

[15] Not all of them: people do make mistakes and tell lies.

[16] Not *all* of them. I suppose that 'The car that left the scene of the accident was *not* a dark green Lexus' could be true even if color were an illusion – but comparatively few of our real-life "color statements" are negative statements.

are to be truly free.[17] We can't conclude from this, however, that then we don't have free will, because free will in this bonkers sense is not the only concept of free will.

And he says this about the latter, early on in the chapter:

> The intuition pumps in this [chapter] are designed to wean you from [the folk ideology of] free will and get you to see a better concept, the concept of real free will, practical free will, the phenomenon in the manifest image that matters.

Unfortunately, Dennett never gets round to spelling out the precise content of either concept (maybe he can't be blamed for that in the case of the folk concept, supposing it to exist; maybe it *has* no precise content). He does tell us quite a bit about both concepts, however. For example, he quotes statements by various philosophers that are intended to illustrate the folk concept – such as this passage from Jerry Fodor's review of his book *Freedom Evolves*: "One wants to be what tradition has it that Eve was when she bit the apple. Perfectly free to do otherwise. So perfectly free, in fact, that even God couldn't tell which way she'd jump." And this passage, from Galen Strawson's review of the same book: "[Dennett] doesn't establish the kind of absolute free will and moral responsibility that most people want to believe in."

Dennett supposes that this "being perfectly free to do otherwise," this "absolute free will" are not philosophers' inventions but are components of the folk ideology of free will – that Fodor and Strawson have correctly (if rather sketchily) identified as what "one wants to be" and what "most people want to believe in." He seems to come down rather hard on Fodor's "even God couldn't tell which way she'd jump" idea – that is, he seems to suppose that absolute unpredictability, unpredictability even in principle, of (some?) human behavior is the essential core of the folk-ideological concept of absolute free will. And I think that he sees Strawson's well-known contention that to enjoy absolute free will one would have to be the sole and ultimate cause of one's actions as also being essential to the folk ideology. He wonders why anyone would want to be an in-principle-unpredictable

---

[17] Whatever "many folk" may think, this is not what philosophers who profess and call themselves 'incompatibilists' think. According to incompatibilism, if an agent decides to do A rather than B, then – if the agent was *able* to choose to do B – *neither* a decision to do A *nor* a decision to do B would have violated the laws of physics and chemistry. In Dennett's defense however, it should be noted that he elsewhere mentions an unnamed philosopher who "has frankly announced that a free choice is a 'little miracle'." As President-for-Life of the World Society of Incompatibilists, it is my unpleasant duty to inform that philosopher that if he or she is a member of the Society, he or she is hereby excommunicated for having made this heretical statement. (The heresy in question is sometimes called 'contra-causal freedom'.)

agent and the ultimate cause of one's actions, and why anyone would be attracted to the belief that we were agents of this sort. And well he might. I would too – at least if I were sure I understood what Fodor and Strawson were talking about. In my view, however, Fodor and Strawson are simply reproducing some ideas invented by philosophers and not reporting what "the man on the Clapham omnibus" wants to be or believes in the existence of. Or, better, not *ideas* invented by philosophers but words and phrases invented by philosophers – "a certain special, happy style of blinkering philosophical English."

In any case, he wants to replace the folk-ideological concept of free will with something else: a better candidate for the office "free will," something that is consistent with our present scientific knowledge, something that isn't bonkers, something that it would actually make sense to want to have oneself and to want to believe that one's fellow human agents had.

My only problem with this project is that there's no such office. Whatever the replacement he may have in mind may be, there's nothing for it to replace.

It's not entirely clear what the proposed otiose replacement is, although it certainly has these features:

> It involves a certain amount of unpredictability in practice, but not unpredictability in principle. (Like it or not, life occasionally requires us to compete with other inhabitants of the world, and organisms that are deficient in unpredictability tend not to pass their genes along to their descendants – think of a gazelle that always swerves left when it's being pursued by a lion. But since we're unlikely ever to be in competition with the Laplacian Reckoner, unpredictability in principle would enjoy no advantage over unpredictability in practice – unpredictability by the organisms with which we are actually in competition.)

> It involves certain kinds of "freedom from" – from coercion, from physical bondage, from illusion and hallucination …

> It is compatible with determinism.

Well, I'm happy to give Dennett the words 'free will'. Let him spell out the details of the concept he intends this phrase to denote as he wants. Possibly he would spell them out the way Liam Clegg does in his paper "Protean Free Will,"[18] a paper for which Dennett (in the chapter of *Intuition Pumps* that I've been examining) has expressed great admiration:

[18] Available on line at http://authors.library.caltech.edu/29887/

Consider an agent who faces an environment which includes sophisticated other agents with interests contrary to hers. Call the agent Mary, and call the other agents predators. One good way for Mary to avoid exploitation by predators, exploitation which may include death, is to engage in *protean behavior* … That is, she may behave somewhat erratically so as to be unpredictable. As documented by Miller … the protean strategy offers many clear advantages over the alternatives of concealment of intentions and active deception. While most notions of 'reason' prescribe a single optimal action in any situation, Mary's behavior must sometimes be locally suboptimal for the sake of unpredictability.

Such local sub-optimality means that at some times t, there are multiple courses of action A_t available which are tied for the strategic optimum. Call the set of such courses of action Mary's strategy S_t. For the purpose of this discussion, no generality is lost in assuming that the optimal probability distribution over S_t is uniform, so that each A_t is equiprobable. Furthermore, while S_t is defined from a bird's-eye view, we may assume that Mary has evolved some reasonably good mechanism for approximating it in the real world. The elements of S_t depend on Mary's preferences, of course, but the strategic optimality of each course of action also includes the risk of exploitation by predators. This, in turn, depends on Mary's past actions and predators' resulting guesses about her next action. If Mary and her predators both use optimal mixed strategies … each A_t should be equally optimal for Mary, and there should be no advantage to a representative predator of predicting that Mary will perform any given $A_t \in S_t$ rather than any other $A'_t \in S_t$. However, if a predator knew *or reasonably suspected* at time r, $r < t$, that Mary were going to perform A^*_t at time t, the predator could exploit this knowledge. A^*_t would therefore no longer be optimal for Mary, and would therefore not be an element of S_t when time t arrived. Call the ability to select an A^*_t from S_t and perform it such that A^*_t is still in S_t at time t 'protean free will' (PFW).

In any case, there's *some* concept Dennett wants to replace the supposed folk-ideological concept of free will with – and it seems to be *something* like Clegg's "protean free will." Let's suppose we have Dennett's concept before us. Whatever precisely it may be, I am, as I have said, happy to let him call it 'free will' – or for that matter, 'the one possible non-bonkers concept of free will' if that's what he wants to call it. And then let him and Fodor and Strawson – those guys on the other side of the verbally essentialist coin – fight over the words 'free will'. Let them fight over who's got free will *right* or over what free will *really is* (or really would be if it existed). Let Fodor and Strawson accuse Dennett of purveying (in Kant's words)

ein elender Behelf – "a miserable substitute" for *true* free will.[19] Let Dennett reply that a belief in the phenomenon Fodor and Strawson want the words 'free will' to designate would be as bonkers as a belief in levitation (a comparison he uses at one point).

I think such a debate would be about an entirely meaningless issue. It's certainly about a meaningless issue if, as I suppose, there is no concept that goes by the name 'free will'. But suppose I'm wrong about that. It's at least not clear that the debate would be a meaningful one even in that case. Suppose that most people *do* believe or want to believe that what they do is in principle unpredictable and *do* believe that they are the sole causes of what they do; why would what Dennett offers as a substitute for what they want or believe in or want to believe in *be* a substitute – even a disappointing substitute – for *those* things? After all, supposing that what Fodor and Strawson say people want is non-existent, still, not just any existent thing counts as substitute for just any non-existent thing, if I may so express myself. ("I want to find the fountain of youth." "Oh, you're like Ponce De León. You want to find the unreconstructed, folk-ideological fountain of youth. That's bonkers. There's no such thing. But, fortunately, the fountain of youth exists; it just isn't what the folk think it is. It isn't a fountain that, when one drinks from it, one is magically restored to youth and then never ages. It's a regimen of diet and exercise that can extend the vigor of one's youth by as much as fifteen years.") The only thing that ties Dennett's substitute to the Fodor–Strawson original – whether that original is a philosopher's invention or an actual folk ideology – is that he calls it 'free will'. In his defense, I'll concede that he has as much right to call some concept of his devising 'free will' as Fodor has to say that a person whose behavior is not predictable even in principle is 'perfectly free to do otherwise' or as Strawson has to call the ability to create one's character *ex nihilo* 'absolute free will'. Since none of these terms – 'free will', 'perfectly free to do otherwise', 'absolute free will' – means anything in particular, they're available to be put to any use a writer wants to put them to.

But – again – suppose I'm wrong. Suppose the idea of a debate about "what free will really is" (or about "what free will should be" or about "the proper meaning of 'free will'") makes sense. Suppose, even, that such a debate would be a philosophically important debate. That meaningful and philosophically important debate, I contend, would be irrelevant to the Culpability Problem.

[19] *Critique of Practical Reason*, I, III, "Critical Examination of the Analytic of Pure Practical Reason." The famous phrase "a wretched *subterfuge*" is Abbott's mistranslation of this description.