CHAPTER 14

Ability

1 The "Classical" Understanding of the Problem of Free Will and Determinism

From the mid 1960s to the mid 1980s the problem of free will and determinism seemed to be a very straightforward problem. Easy to solve (and solved) according to some, difficult to solve (and unsolved) according to others, difficult to say anything new about according to most, but straightforward in the sense that everyone (that is, every analytical philosopher) knew what the problem *was* – everyone agreed, within very broad limits, about how the problem should be stated or posed. This is not to say that a philosopher writing in (say) 1968 would have set out the problem in exactly the same way as a philosopher writing in (say) 1982. Harry Frankfurt's essay "Alternate Possibilities and Moral Responsibility"[1] was responsible for an important change in the way the problem was formulated (a point to which I shall return).[2] But at any given moment in the classical era – as I shall call the mid 1960s to the mid 1980s – everyone accepted formulations of the problem that were more or less the same.

How did philosophers of the classical era see the problem of free will and determinism?

It was agreed, first of all, that the phrase 'free will' was not to be taken seriously. Everyone conceded that 'free will' had become something like a proper name: like 'the Holy Roman Empire', it was what something had once been called for what had then seemed to be a good reason and what it was now called for no better reason than that it had been called by that name for a very long time.

And everyone during the classical era agreed that this thing inappropriately called 'free will' was a sort of *n*-way power with respect to the future. Agents often deliberate between two or more courses of action, and to

[1] *Journal of Philosophy* 66 (1969): 829–839.

[2] "Difficult to say anything new about" – but not impossible, for in that essay Frankfurt *did*.

ascribe *free will* to a rational agent (everyone agreed) was to say that at least sometimes when that agent is trying to decide what to do, it is within that agent's power to do *each* of the things he or she is trying to decide between. Not *all*, mind you, but *each*. Suppose, for example, that Sally, who is in a sticky legal situation, is trying to decide whether to lie about where she had been on the night of March 11th or to tell the truth or to remain stubbornly silent. It may well be within her power to lie *and* within her power to tell the truth *and* within her power to remain silent – but it obviously could not be within her power to do all three (or any two) of those things. (Of course, she might vacillate – she might lie and then change her mind and tell the truth, for example. But in the end she is going to have to choose exactly one of the three options facing her.)

The concept of a course of action being "within one's power," it was universally held in the classical era, was to be explained by an appeal to one of the several meanings of the ordinary, everyday word 'can'. To say that Sally has the three-way power to lie or to tell the truth or to remain silent, for example, is to say simply that if she were to say these three things to herself

I can lie

I can tell the truth

I can remain silent

all three statements would be true. And these statements (a philosopher of the classical era would have said) are verbal variants on

I am able to lie

I am able to tell the truth

I am able to remain silent

and

It is within my power to lie

It is within my power to tell the truth

It is within my power to remain silent.

It was generally recognized in the classical era that the concept expressed in such contexts as this by 'can' and 'able' and 'within one's power', while certainly in some sense a modal concept, the concept of a certain sort of possibility, is not the modal concept that goes by the name 'possibility' in modal logic. For no sense of 'it is possible that' are the statements

Sally can lie

Sally is able to lie

It is within Sally's power to lie

equivalent to the statement

It is possible that Sally will lie.[3]

They are not, for example, equivalent to 'It is physically possible that Sally will lie'. *Some* philosophers of the classical era, it is true, would have accepted the thesis that (e.g.) 'Sally is able to lie *only if* it is physically possible that Sally will lie' was a conceptual truth, but no one would have accepted the thesis that 'Sally is able to lie *if* it is physically possible that Sally will lie' was a conceptual truth.[4] The reason was simple enough. If it is physically possible that Sally will lie, that means that there is a "physically possible future of the world" (a future of the world permitted by the present state of the world and the laws of physics) in which Sally lies.[5] But from the fact that such a possible future exists, it does not follow that it is within Sally's power to ensure or see to it or bring it about that that possible future will be a future that actually comes to pass.

Suppose, for example, that I am locked in a room and that I am unable to unlock the door. I wish to leave the room – and I *shall* leave the room if something unlocks the door. As matters stand, however, the only "something" that can unlock the door is an indeterministic mechanism built into the door: there are physically possible futures in which the mechanism unlocks the door and there are physically possible futures in which it does not – and in some of those in which it does, I leave the room. But in the actual future[6] it is not going to unlock the door. (And therefore, I am unable to leave the room.)

[3] But some would have held that these three statements were equivalent to 'It is possible *for* Sally *to* lie'.

[4] Carl Ginet might be thought to be an exception to this generalization, but this would not be right. Ginet would certainly accept (and would have accepted during the classical period) the counterexample that I shall present in the next paragraph of the text.

[5] At any rate, that is one of the things one might mean by 'physical possibility'. There is another and weaker sense of the phrase: something is physically possible if it occurs in some world in which the laws are the same as those of the actual world (whether the past is the same as that of the actual world or not). That weaker sense obviously does not imply ability, for in the weaker sense it is physically possible that I, who am at the moment in North America, now be in Sydney.

[6] This argument presupposes that statements "about the future" all have truth-values – and that there is therefore exactly one "actual" future, to wit, the future such that a statement about the future is true if and only if it is true in or according to that future. In the example we are considering, it is undetermined by the laws of physics and the present state of things whether the mechanism will unlock the door; but the mechanism *will not* in fact unlock the door: that in fact *is* how the indeterministic evolution of the world is going to proceed.

Suppose we say that a sentential operator M is a *possibility operator* if, for some set of possible worlds S and any declarative sentence p, the sentence $\ulcorner \mathrm{M}(p) \urcorner$ is true just in the case that p is true in some member of S.[7] Anthony Kenny showed (by an argument much more abstract than the argument of the preceding paragraph) that for no possibility operator M are sentences of the forms

X can ϕ

X is able to ϕ

It is within *X*'s power to ϕ

equivalent to the corresponding sentence of the form

M (*X* ϕs).[8]

The essential point of Kenny's argument may be stated as follows. Suppose that if one ϕs then one may or may not χ and one may or may not ψ, but if one ϕs one must either χ or ψ. (For example, if one throws a dart that hits the dartboard, one may or may not throw a dart that hits the left-hand side of the board and one may or may not throw a dart that hits the right-hand side of the board, but, if one throws a dart that hits the dartboard, one must either throw a dart that hits the left-hand side of the board or throw a dart that hits the right-hand side of the board. If one arrives in Chicago, one may or may not arrive on a Tuesday and one may or may not arrive on some other day of the week, but one must either arrive on a Tuesday or arrive on some other day of the week.)

Now for any possibility operator M, and any sentences p and q,

$\ulcorner \mathrm{M}(p \vee q) \rightarrow .\mathrm{M}(p) \vee \mathrm{M}(q) \urcorner$

is a valid sentence (owing to the fact that, if S is the relevant set of worlds, $\ulcorner \mathrm{M}(p) \urcorner$ is true if p is true in some member of S; and if a disjunction is true in some member of a set of worlds, one at least of its disjuncts must be true in that world). And, therefore, for no possibility operator M is

X is able to ϕ

equivalent to

M(*X*ϕs).

[7] *S* might be, for example, the set of all worlds; or it might be the set of worlds in which the laws of nature are the same as those of the actual world; or it might be the set of worlds in which the laws *and* the present state of things are the same as they are in the actual world.

[8] See Anthony Kenny, *Will, Freedom and Power* (Oxford: Blackwell, 1975), pp. 137–140.

For if this equivalence holds, and if ϕ, χ and ψ are related in the way set out above, '$X\,\phi$s' is true in the same worlds as '$X\,\chi$s $\vee$ $X\,\psi$s' and

X is able to $\chi \vee X$ is able to ψ

is therefore strictly implied by

X is able to ϕ.

But the latter may be true and the former false. For it may be true that I am able to throw a dart that will hit the dartboard but false that I am able to throw a dart that will hit the right-hand side of the board *and* false that I am able to throw a dart that will hit the left-hand side of the board. I may be able to ensure that I arrive in Chicago but have no choice whatever as to what day of the week I arrive in Chicago – the airlines being what they are these days.[9]

Finally, a word about the phrase (much used in the classical era) 'could have done otherwise'. Let us again consider Sally, who is deliberating about whether to lie or to tell the truth or to remain silent. Suppose that she *can* lie and *can* tell the truth and *can* remain silent. Suppose further that time passes and what she *does* do is lie. Then, since 'could have' is the present perfect tense of the modal auxiliary verb 'can', we can say in retrospect that she *could have* told the truth and *could have* lied and *could have* remained silent[10] – which, of course implies that she *could have done otherwise*. '*X* could have done otherwise' is nothing more than a retrospective, present perfect "version" of the present tense statement '*X* must choose among two

[9] Suppose someone insists that if I do something, then, before I did it, I must have been *able* to do it – that if, for example, I (who have no skill at manipulating playing cards) draw a card at random from a standard deck and it is the four of clubs, then it follows that, before I drew the card, I could have said truly, "I am able to draw the four of clubs." As we shall see in Section 2, there are many senses of 'able'. Let us say that any of them that has this feature is "minimal" (a more or less arbitrarily chosen word). Any sense of 'able' in which only a card-sharp or a stage magician is "able" to draw the four of clubs (apparently at random) from a standard deck is therefore not minimal. I will concede for the sake of avoiding a merely verbal dispute that there are minimal senses of 'able' (although I doubt whether there are). But I insist that if there are minimal senses of 'able', there are certainly non-minimal senses as well. All occurrences of 'able' in the remainder of Section 1 should be understood as having some non-minimal sense. (Similar remarks apply to 'can'.)

[10] The simple past tense of 'can' is 'could' ('I could speak French when I was a child, but I can't now'). But the simple past 'could' seems to be capable of expressing only the "skill" or "general" sense of ability (see note 29 below). It is therefore necessary to use the present perfect 'could have' to make statements about our past abilities with respect to, as one might say, particular occasions. Thus a woman might say 'I could have told my husband the truth yesterday' but not 'I could tell my husband the truth yesterday'. ('I could tell my husband the truth yesterday' would probably strike most speakers as a very puzzling statement, although one might make some sort of sense of it by assimilating it to such "longer-term" general-ability statements as 'I could tell my husband the truth about all my activities before I joined the CIA, but now I have to lie to him about practically everything'.)

or more alternatives and can (i.e., is able to, has it within her power to) choose each of them'.

If we call the preceding paragraphs – from 'The concept of a course of action …' to this point – "an account of ability," that account of ability is (more or less) the account of ability that was presented in the course Philosophy 301, The Problem of Free Will and Determinism (or whatever it might have been called), in the classical era.[11]

I have mentioned Frankfurt's remarkable essay "Alternate Possibilities and Moral Responsibility." The classical era may usefully be regarded as falling into two parts – before and after the publication of that essay. Or perhaps it would be more accurate to say "before and after the philosophers began to appreciate the power and significance of the arguments of 'Alternate Possibilities and Moral Responsibility'," and this did not happen till at least four or five years after the essay was published.

In the pre-Frankfurt years, so to call the part of the classical era before the arguments of Frankfurt's essay began to be widely known and appreciated, philosophers' understanding of "the problem of free will and determinism" can be described as follows. (So, at any rate, I contend.)

Suppose, first, that we say that 'determinism' is the thesis that the laws of nature (or the laws of physics) and the state of the physical world at any given moment determine the state of the physical world at any other moment. Then, in the pre-Frankfurt years, work on "the problem of free will and determinism" consisted almost entirely in attempts to provide and defend answers to the following two questions:[12]

Question 1 Is the existence of free will compatible with determinism?

Question 2 Is the existence of free will compatible with indeterminism?[13]

(Indeterminism, of course, is simply the logical contradictory of determinism.) These two questions were held to be important because everyone[14] accepted the thesis

> Moral responsibility requires the existence of free will.

[11] Kenny's argument was published too late to have been known throughout the entire classical era, and was never as well known as it should have been. But the proposition that was to be its conclusion was well known and generally accepted throughout the classical era, owing to the influence of Richard Taylor's classic essay "I Can" (*Philosophical Review* 69 [1960]: 78–89).

[12] Or, perhaps, of these two questions and such further questions as might be raised by the various answers proposed for them.

[13] Actually, Question 2 needs to be expressed more carefully than this. The necessary qualifications will be stated later in the text (and in note 16).

[14] When I say that "everyone" – the universe of discourse being analytical philosophers who wrote on the problem of free will – accepted a certain view, I mean 'practically everyone' or 'more or less everyone' or 'pretty much everyone'.

Everyone agreed that moral responsibility was an important thing. And, therefore, everyone agreed that free will was an important thing – if only because it was necessary for moral responsibility. (Some philosophers thought that free will might be important for other reasons as well – it was, for example, commonly held to be a valuable thing in itself.) And, of course, if free will exists, it must be compatible with determinism or (inclusive) compatible with indeterminism. But there were important arguments for both the incompatibility of free will and determinism and the incompatibility of free will and indeterminism – hence, the importance of the two questions. It was commonly, if tacitly, held that if these two questions could be given satisfactory answers, the problem of free will and determinism could be regarded as solved – for those questions constituted the hard or philosophically interesting part of the problem of free will and determinism.

Most philosophers of the pre-Frankfurt years accepted one of the following two positions:

1. Determinism is true;[15] free will exists.

2. The existence of free will is incompatible with the truth of determinism; free will exists.

Position (1) was called (following William James) "soft determinism." Soft determinism immediately implies *compatibilism*, the thesis that the existence of free will is compatible with the truth of determinism. Position (2) was called 'libertarianism'. Its first component, the thesis that the existence of free will is incompatible with the truth of determinism, was called *incompatibilism.* Debates about the problem of free will and determinism in the pre-Frankfurt era, therefore, were largely debates about whether free will was compatible with determinism. (The soft-determinist defenders of compatibilism were far more numerous than the libertarian defenders of incompatibilism.) That is to say, the single most frequently addressed question in discussions of the free will problem was Question 1. But Question 2 was not neglected, owing to the fact that many, perhaps most, soft determinists accepted the following thesis (which is not logically implied by soft determinism):

> The existence of free will is not only compatible with the truth of determinism, it is *incompatible* with the truth of *indeterminism* – or if

[15] Said with some sort of bow to quantum mechanics, something along the lines of 'Quantum mechanics apparently implies that some aspects of the "micro-world" are indeterministic; nevertheless, the "macro-world" is deterministic – or as nearly deterministic as makes no matter'.

> not with the truth of indeterminism "in general," at any rate with the truth of the thesis that *human acts* are undetermined.[16] For if human acts were undetermined, they would be "bolts from the blue," events that were not grounded in the beliefs, values, or character of their agents, mere *intrusions* into the lives of their agents.

The position that James had called 'hard determinism',

> Determinism is true; the existence of free will is incompatible with its truth,

a position held by most of the eighteenth- and nineteenth-century materialists, was held by hardly any analytical philosopher in the pre-Frankfurt years, for in those years everyone believed in free will. One had to, you see, because free will was required for moral responsibility – and, in particular, moral responsibility for wrong-doing.[17] Obviously (everyone supposed) some people have behaved in ways in which they ought not to have behaved, and one cannot be responsible for having done something one ought not to have done if one was not able to do otherwise. If you had asked a philosopher of the pre-Frankfurt years why this was so, the response would have been something along the following lines.

Suppose you tell someone who has done *X* that he ought not to have done *X*. That statement presupposes that he ought to have done something other than *X* (let "doing nothing at all" be a special case of "doing something other than *X*"). You are in effect saying to him, "You ought to have done something other than *X*." But if you tell someone that he ought to have done something, by the very fact of making that assertion you

[16] This statement contains the qualification of Question 1 that was promised in note 13 above. Suppose that Sally, who had been deliberating between lying and telling the truth, lied at t_1. Her lying at t_1 was an undetermined act just in the case that there was a moment t_0 shortly before t_1 such that in some of the futures allowed by the state of the world at t_0 and the laws of nature she lied *and* in some of those futures she did not lie. A definition of 'undetermined act' would be a generalization of this statement about Sally's act of lying. Let us say that an act that is not undetermined is determined. It follows from determinism that all human acts are determined. It does not follow from indeterminism that any human act is undetermined. If, for example, there is one particle in intergalactic space whose behavior is undetermined (and whose behavior affects nothing else) and if the behavior of everything else is fully determined, indeterminism is true and all human acts are determined.

[17] The eighteenth- and nineteenth-century materialists would have agreed. But those extremely "tough-minded" philosophers (to borrow another of James's pithy phrases) were willing to say that moral responsibility was as much an illusion as the free will whose existence it presupposed. (Some of them at any rate: those who, like Holbach, were incompatibilists – not, of course, a word any of them would have known. But some would have followed their seventeenth-century predecessor Thomas Hobbes and embraced compatibilism.)

commit yourself to accepting the thesis that he was *able* to do that thing. You can't, for example, tell King Canute that he ought to have halted the advance of the tides (and mean it) unless you believe that he was able to halt the advance of the tides. Hard determinism therefore implies that if a statement of the form 'You ought not to have done *X*' is addressed to someone who has done *X*, that statement *must* be false – or at least embody a false presupposition. Without committing ourselves to any controversial position on the semantics of moral judgments, we can say that hard determinism implies that any such statement will be a defective statement for the same reason (whatever it may be) as 'You ought to have halted the advance of the tides' would be a defective statement if it were addressed to Canute.

Enter Frankfurt. This is "the Principle of Alternate Possibilities"[18] – PAP for short – that was the topic of the essay of that name:

> A person is morally responsible for what he has done only if he could have done otherwise.

Frankfurt presented convincing counterexamples to this principle. The essential idea behind these counterexamples is the idea of an "offstage counterfactual manipulator." This is a typical "Frankfurt counterexample" to PAP:

Poisson has put arsenic in Dyer's tea, and this action caused her very soon afterwards to die of massive organ failure. Add to this case whatever *you* think is needed to make 'Poisson is morally responsible[19] for having poisoned Dyer' true: Poisson acted with the intention of causing her death, Poisson was sane and not of subnormal intelligence, Poisson was able to refrain from poisoning her, Poisson's poisoning her tea was undetermined … whatever you like. So: Poisson poisoned Dyer and is responsible for having done so. Now expand the story of Poisson and Dyer as follows. There was an evil genius, Manipula, who, for reasons of her own, in the days leading up to Dyer's death very much wanted Poisson to act on his intention to poison Dyer. Manipula could see, by looking into Poisson's soul, that he firmly intended to poison Dyer. And she could see that he had laid his plans very carefully. She therefore thought it almost certain that he would carry them out – almost certain, but not *perfectly* certain. Manipula was not the sort of evil genius to leave anything to chance, and she accordingly devised the following "contingency plan."

[18] Of course it ought to have been called 'the Principle of Alternative Possibilities'.
[19] In the sequel, 'responsible' will mean 'morally responsible'.

If Poisson shows any sign of wavering in his present determination to poison Dyer, I will – by direct manipulation of his brain – suppress all doubts and reservations, moral or practical, that may have crept into his mind, fill his mind with a sense of absolute certainly that his plan will succeed and that he will never be so much as suspected of poisoning Dyer. And, finally, I will strengthen his desire to poison her till it is irresistible.

Manipula had the power to do all the things she had (conditionally) decided to do. (And she was of such a nature that – unlike a human being – when she had made up her mind to do something, it was impossible for her to change it.) In the event, however, Poisson never "wavered": he went ahead and poisoned Dyer just as he had resolved to do, and Manipula's contingency plan never had to be put into effect. That is to say, she did *nothing* that affected Poisson in any way. It is evident, therefore, that what we have *added* to the "original" story of Poisson and Dyer does not remove, diminish, undermine, militate against – choose what verb you will – Poisson's responsibility for having poisoned Dyer. Therefore, since Poisson was responsible for that act in the original story, he was responsible for it in the expanded story, the story incorporating Manipula and her contingency plan.[20] But note: in the expanded story, Poisson was unable *not* to poison Dyer. Manipula's forming her contingency plan had the effect of "pinching off" all those possible futures (if there were any) that commenced shortly before the moment at which (in actuality) he poisoned Dyer and in which he changed his mind and did not poison her. *All* the possible futures that confronted Poisson as he slipped the packet of arsenic trioxide into his waistcoat pocket and set out for Dyer's house on that fatal day were futures in which he was going to poison her. Manipula's purely counterfactual plan, her *unacted-on* plan, has rendered non-existent all the "alternative possibilities" (alternative, that is, to his poisoning Dyer) that existed in the original story. PAP is therefore false – or at any rate, not a conceptual truth.

Most philosophers working on the problem of free will and determinism found Frankfurt's arguments convincing.[21] As a result, a third "general position" as regards the problem of free will and determinism emerged (in addition to soft determinism and libertarianism). Following John Martin

[20] Of course, if the contingency plan *had* been put into effect, then no doubt Poisson *would not* have been responsible for poisoning Dyer. But it wasn't. And he was.

[21] For my own views on the import of "Frankfurt counterexamples," see Chapters 1 and 6 of the present volume.

Fischer, we may call it "semi-compatibilism" (although this phrase was coined well after the close of the classical era):[22]

> Free will may or may not be compatible with determinism, but moral responsibility is compatible with determinism.

Semi-compatibilists, understandably, lost interest in Question 1 (and in Question 2). For, as I said above, 'These two questions had been held to be important because everyone had accepted the thesis "Moral responsibility requires the existence of free will."' And this was precisely the thesis that semi-compatibilists did not accept.

This must suffice for an account of how the problem of free will and determinism was understood in "the classical era." It is far from being a complete account. (I have, for example, said almost nothing about contemporary discussions of Question 2 – and nothing at all about the debates about "agent causation" that arose out of that discussion.) In my view, the classical era had it right.[23] The problem was correctly formulated, and philosophers discussing the problem of free will and determinism should focus on three questions: Question 1, Question 2, and

Do Frankfurt's arguments indeed show that ascriptions of moral responsibility do not imply the existence of free will (do not imply that human beings are sometimes able to act otherwise than they in fact do)?

Now suppose that, like me, you think that the answer to the third question is No: whatever the value of, whatever the import of, Frankfurt's arguments, whatever may be right about them, nevertheless, if no one is ever been able to do otherwise than he or she in fact does, then no one is ever morally responsible for anything.[24] Then you will believe, as I do, that Question 1 and Question 2 are the central questions of the problem of free will and determinism, and you will think that the problem of free will and determinism is an important problem (if for no other reason) because moral responsibility is important and moral responsibility is impossible if the ability to act otherwise than one does is incompatible both with one's actions being determined and with one's actions being undetermined.

[22] John Martin Fischer, *The Metaphysics of Free Will* (Oxford: Blackwell, 1994), p. 178ff.

[23] I am extremely skeptical about the value of almost all work done on the problem of free will and determinism (or the problem of moral responsibility and determinism) after the classical era. For my reasons for this, see Chapters 10 and 13 of the present volume.

[24] See Chapter 1 of the present volume. For a condensed version of the argument of this chapter, see Peter van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983), pp. 162–182.

And, if you believe these things, you will be interested in the question: What is ability – what is the meaning of the word 'able' in phrases like 'able to lie and able to tell the truth' and 'able to act otherwise' as these phrases were used in the discussions of the problem of free will and determinism in the classical era?

2 The Concept of Ability in the Classical Understanding of the Problem of Free Will and Determinism

The meaning of the word 'able' as this word was used in discussions of the problem of free will and determinism in the classical era, is, I believe, best explained in connection with a certain way in which promises can be defective. To specify the "way" I have in mind, it will be necessary to contrast it with other ways in which a promise can be defective. (And I hope that these examples will also make it clear what I mean by describing certain promises as 'defective'.) We shall examine some imaginary cases in which a promise is made and that promise is in one way or another defective.

A defective promise – case 1 Mr. Rich, the prominent industrialist, is about to have a delicate brain operation. His wife presses Dr. Sturgeon, the Chief of Surgery at St Luke's Hospital, to promise her that the operation will be performed by the most skilled brain surgeon on his staff. Sturgeon promises her this, knowing that if he did not make that promise she would have her husband's surgery performed in another hospital. And that is exactly what Sturgeon does not want, for he is both a man of the left and a strict utilitarian and believes that the overall utility of the world would be increased significantly if Rich, a powerful reactionary, were to die on the operating table. He further believes that if Rich remains at St Luke's, he can render that outcome reasonably probable – for he plans to assign the operation to a brain surgeon called Sharp whom he privately, and with some justification, believes to be a blundering incompetent. "If anyone can kill the bastard, it's old 'Notso' Sharp," Sturgeon says to himself.

This is a very straightforward kind of defective promise: Sturgeon promised to do something that he did not intend to do – that is, that he intended not to do. Following Mele, we may say such a promise is defective because it fails to be *sincere*[25] – or, more briefly, is *insincere.* And I suppose we also could say that a promise is insincere if the person making the promise has

[25] Alfred R. Mele, "Agents' Abilities," *Noûs* 37 (2003): 447–470; see p. 453.

not yet decided whether to keep it. (Suppose Sturgeon is *toying* with the idea of assigning the operation to the incompetent Sharp, but has certain moral qualms about doing this even when the patient is someone like Rich – he's not a strict utilitarian as he was in the first version of the story. He promises to assign his best surgeon to the operation in order to keep the "have Sharp do it" option open while he ponders the morality of the matter further.)

A defective promise – case 2 Mr. Rich is about to have an operation. The operation has very little chance of success – it's all but hopeless. His wife presses Dr. Sturgeon (who is to perform the operation) to assure her that the operation will be a success. (She doesn't know that the operation has little chance of success, for our story takes place in the bad old days when doctors never told patients bad news and never told the patients' relatives bad news till the patients were actually dead. Thank God those days are past!) Sturgeon has nothing against Mr. Rich, and would save him if he could, but he knows he can't[26] – or at any rate that his chances of doing so are very, very slim. Now Sturgeon hates emotional scenes, so he says, "I promise you, Mrs. Rich, that the operation will be a success and your husband will be completely restored to health."[27] (He intends to leave the hospital after the operation by a route that will enable him to avoid an encounter with Mr. Rich's widow.)

This is not a case of someone's promising to do something he or she intends not to do. And it's not a case of promising to do something that – one is aware – one might *later* decide not to do. It might well be called an insincere promise – there's certainly something insincere about it – but it's not insincere in the sense in which the promise in case 1 (or the variant on case 1 mentioned in the parenthesis) was insincere. Nevertheless, it seems clear that something is seriously wrong with Sturgeon's promise. It is, as I have said, defective. What is its defect? Does it lie simply in the fact that the person making the promise is unable to keep it (or almost certainly unable to keep it)? No, for there are non-defective promises that the person making the promise is unable to keep. For example:

> **A non-defective promise** Mr. Rich is about to have an operation. The operation has very little chance of success – it's all but

[26] Perhaps he "can" (and hence doesn't know that he can't) in what we earlier called a minimal sense of ability (see note 9 above). If, in the event, a medical "miracle" occurs and the operation is a success, then if Sturgeon had said beforehand, "I can [minimal sense] save the patient," he would have spoken truly.

[27] Or, if you like, "I promise to restore your husband to health." One might dispute about whether a "promise that" [something will be the case] is a promise in the same sense as the sense in which a "promise to" [do something] is a promise.

hopeless. But this fact has to do with some peculiarities of Mr. Rich's brain that are unknown to Dr. Sturgeon (who is to perform the operation); Sturgeon is certain (and given the evidence he has, justifiably certain) that the operation will be a simple and straightforward one, as simple and straightforward as the most routine appendectomy. Mrs. Rich presses Sturgeon to assure her that the operation will be a success. Sturgeon says, "I promise you, Mrs. Rich, that the operation will be a success and your husband will be completely restored to health." (He is shocked and chagrined when bizarre and unforeseeable complications – a consequence of Mr. Rich's aforementioned physiological peculiarities – transpire during the operation, and Mr. Rich dies on the table.)

Contrast that case with the following case:

A defective promise – case 3 Mr. Rich is about to have an operation. The operation will be a simple and straightforward one, as simple and straightforward as the most routine appendectomy – and, for that reason, it will be a success. But these facts are unknown to Dr. Sturgeon (who is to perform the operation); owing to some bizarre flaw in the evidence available to him, Sturgeon is certain (and given the evidence he has, justifiably certain) that the operation has very little chance of success – that it's all but hopeless. Mrs. Rich presses Sturgeon to assure her that the operation will be a success. Sturgeon has nothing against Mr. Rich, and would prefer to save him, but he firmly believes that he (almost certainly) can't. Now Sturgeon hates emotional scenes, so he says, "I promise you, Mrs. Rich, that the operation will be a success and your husband will be completely restored to health." (He intends to leave the hospital after the operation by a route that will enable him to avoid an encounter with Mrs. Rich – who, he now believes, will then almost certainly be Mr. Rich's widow.)

It seems evident from consideration of case 2 and "A non-defective promise" and case 3, that, although the first of the following two principles is false, the second is true.

(F) A promise is necessarily defective if the person making the promise is unable to keep it

(T) A promise is necessarily defective if the person making the promise believes that he or she is unable to keep it.

Now let us consider a variant on case 2:

> **A defective promise – case 2′** Mr. Rich is about to have a delicate brain operation. His wife presses Dr. Sturgeon (who is to perform the operation) to assure her that the operation will be a success. Sturgeon knows that the operation may succeed and that it may fail. The CT scans are inconclusive. Everything is going to depend on what Sturgeon finds when Mr. Rich's skull has been opened and he can visually examine the condition of things in the cranial pia mater. Sturgeon hopes the operation will be a success, but he really has no idea what its outcome will be. Now Sturgeon hates to be in the presence of anyone who is in emotional distress, so he says, "I promise you, Mrs. Rich, that the operation will be a success and your husband will be completely restored to health." (He intends to leave the hospital after the operation by a route that will enable him to avoid an encounter with Mrs. Rich if the operation is a failure.)

It seems clear that case 2′ is case of a defective promise. But it would not be right to say that, when Sturgeon makes the promise, he *believes* that he is unable to do what he is promising to do (that is, unable to bring about the outcome he has promised will occur). Rather, he doesn't know whether he is able to do what he is promising to do. That the promise in case 2′ is defective can be accounted for by the following principle:

(T′) A promise is necessarily defective if the person making the promise does not believe that (does not have the belief that) he or she is able to keep it.

But in what sense of 'able'? Let us continue to assume (I by no means endorse this position) that there are minimal senses of 'able'. Sturgeon certainly does not have the belief that he is able to keep his promise in even a minimal sense. (He may well believe that *for all he knows* he is able to keep the promise in a minimal sense – after all, for all he knows, the operation will be a success. But to have that belief is not to have the belief that he *is* able [minimal sense] to keep the promise.) There are other senses of 'able',[28] but it would seem that there is *no* sense of 'able' in which Sturgeon believes that he is able to keep his promise. (Although it may well be that for *every* sense of 'able' *x* he believes that *for all he knows* he is able to keep the promise in sense *x*.)

[28] For a discussion of the many senses of 'able', see *An Essay on Free Will*, section 1.4 (pp. 8–13). Mele's "Agents' Abilities" presents a rival account – much more fine-grained than mine – of the senses of 'able'. I do not accept all the distinctions that Mele makes.

Let us say that if *x* and *y* are two senses of 'able', *x* is *stronger* than *y* just in the case that (i) if someone is able to do something in sense *x*, then, necessarily, that person is able to do that thing in sense *y*, and (ii) it is possible for there to be someone who is able to do something in sense *y* but is not able to do that thing in sense *x*. For example, a sense of 'able' in which only people with the skills of a card-sharp are able to deal themselves a flush in hearts is stronger than any minimal sense; there's a sense of 'able' in which Grigory Sokolov is able to play Chopin's Prelude in E Minor even when no piano is available to him,[29] and there's a stronger sense in which he's able to play that difficult work only when he has access to a piano; a loan officer who is "able" to approve the loan I've applied for because all she has to do to approve it is to sign her name on a certain piece of paper is not lying when she tells me she's unable to approve it: she has a stronger sense of 'able' in mind.[30] (Say that two senses of 'able', *x* and *y*, are *equivalent* if, necessarily, someone is able to do something in sense *x* if and only if that person is able to do that thing in sense *y*. Note that if two senses of 'able' are non-equivalent, it does not follow from our definition that one of them is stronger than the other.)

I will now define a certain sense of 'able' that, perhaps tendentiously, I will call the Relevant Sense.

> Someone is able in the Relevant Sense (is "able$_{RS}$") to do something just in the case that that person is able to do that thing in the *strongest* sense of 'able' such that, if one made a promise and did not believe that (did not have the belief that) one was able (in that sense) to keep that promise, that promise would be defective.[31]

I have said that in case 2′ "it would seem that there is *no* sense of 'able' in which Sturgeon believes that he is able to keep the promise." I contend that, for every sense of 'able' *x* in which Sturgeon fails to have the belief

[29] This is the '"skill" or "general" sense of ability' sense alluded to in note 10 above.

[30] One sense of '*x* is able to *Y*' is 'If *x* were to choose to *Y*, *x* would *Y*'. Call this the Conditional Sense. Let '**ABLE**' represent any sense of 'able' other than the Conditional Sense itself. 'If *x* were to choose to *Y*, *x* would *Y* – *and x* is **ABLE** to choose to *Y*' is a sense of 'able' that is stronger than the Conditional Sense. This is an adaptation of a point that frequently surfaced in discussions of compatibilism in the classical era, for in those days compatibilists often maintained that to have free will was simply to be able to have done otherwise than one did in the Conditional Sense of 'able'.

[31] But suppose that there are two senses of 'able' that satisfy the condition *vis-à-vis* promising laid down in the definition, that no stronger sense of 'able' does, and that neither sense is stronger than the other (perhaps the two senses are equivalent; or perhaps they are simply "incommensurable": they are not equivalent but neither is stronger than the other)? I'll cross that bridge if I come to it – that is, if someone presents a plausible example of two such senses. The definition presupposes that there *is* a strongest sense of 'able' that satisfies "the condition *vis-à-vis* promising." If there is no such sense, the definition will, of course, have to be revised.

that he is able in sense *x* to keep the promise he has made, one is the strongest: *if* Sturgeon is able to keep his promise in *that* sense, he is able to keep it in every other sense of 'able'. And – surely? – if that sense exists (and if minimal senses of 'able' exist) that sense is stronger than any minimal sense. The definition I have given may be described as a *functional specification* of the "Relevant Sense." It is natural to ask whether an *analytic definition* of the Relevant Sense can be given. (As 'undefeated justified true belief' was once supposed by some to be an analytic definition of 'knowledge'.) I am sorry to have to say that I have none to offer. (It is certainly implausible to suppose that such an analytic definition of 'able$_{RS}$' could be devised if there were no other, better understood sense of 'able' that could figure in its *definiens*.)

Now one might wonder whether I can *justifiably* believe that I am able (in *any* sense of 'able') to keep a promise I have made.[32] Suppose, for example, you ask me to give you a ride to work tomorrow morning and I say that I will (thereby promising to give you a ride to work tomorrow morning). How do I know that I am able to do that? (Perhaps it would be more natural to frame the question in the future tense: 'How do I know that I *shall* be able to do that?" If this is so, it strikes me as a mere matter of idiom. Since the modal auxiliary 'can' has no future tense, one is forced to use the present in the parallel case: "How do I know that I *can* do that?"; there doesn't seem to be anything logically odd about that question.) After all, my car (which has started right up every day for the last three years) *might* not start tomorrow. My hitherto reliable alarm clock *might* fail me. Despite my apparent good health, I *might* die in my sleep. Literally hundreds of things whose non-occurrence I'm not in a position to predict with certainty *might* happen to prevent me from keeping my promise to you. And if I don't know whether I am (or shall be) able to give you a ride to work tomorrow, this is presumably because I am not justified in believing that I am able to give you a ride to work tomorrow. The same point, of course, applies to any conceivable promise. And if I can't justifiably believe that I am able (in any sense) to do a certain thing, and I know that I can't justifiably believe that I have that ability, then I ought *not* to believe that I am able to do that thing. And, therefore, if there is a sense of 'able' such that a promise I make at *t* is defective if at *t* I do not have the belief that I am able to keep it, then either I ought never to make any promises (because doing so would require me to have a belief that is not justified) or all promises are defective (which would also seem to imply that I ought never to make any promises). But

[32] See section 4 (pp. 457–461) of Mele's "Agents' Abilities."

it's evident that (in normal circumstances) I violate none of my epistemic duties in promising to give you a ride to work tomorrow and it's evident that not all promises are defective.

This sort of argument, it seems to me, sets a very high standard for a belief's being justified. Anyone who adheres to this standard ought to say that I should never make any statement about the future. I should never say things like, "I'll see you in Chicago on Thursday" or "At our next meeting we'll discuss Thomson's Trolley Problem." (And of course I should never believe what others say when they make similar assertions about the future.) Well, I'll leave it to the epistemologists to sort this one out. Whatever the correct epistemological account of justified beliefs about the future (beliefs about the future of the simplest, most straightforward sort) should be, it is not very plausible to suppose that it will imply that my beliefs about what I'm going to be doing the day after tomorrow are necessarily unjustified.

So we have the Relevant Sense of ability. And what I am contending it is relevant to is, of course, the classical understanding of the problem of free will and determinism. That is to say, the classical understanding of the problem turns on a definition of free will that is something very much like the following:

> x has free will $=_{df}$ x must sometimes choose among two or more alternative courses of action and, on at least some of these occasions, x is able$_{RS}$ to choose each of them.

I say that the classical understanding of the problem of free will and determinism turns on this definition of free will (or on one very much like it) because it is in this sense of 'free will' that an agent's having free will is, or seems to be, incompatible with the agent's actions being *un*determined.[33] Disputes about the compatibility of free will and universal causal determinism in the classical era were much less sensitive to the precise meaning of 'able' that figured in the disputants' definitions of free will – for the usual arguments for the incompatibility of free will and determinism required only that free will be understood as involving a sense of 'able' (a) such that in that sense no agent is able

– to render false a necessary truth
– to render false a true proposition about the past
– to render false a law of nature,

[33] See Chapter 11 in this volume.

and (b) such that all instances of the following schema are valid if the word 'able' therein is understood in that sense:

> *p* and no one is or ever has been able to render the proposition that *p* false
>
> (If *p* then *q*) and no one is or ever has been able to render the proposition that if *p* then *q* false
>
> *hence*,
>
> q and no one is or ever has been able to render the proposition that q false.[34]

There seems to be *no* sense of 'able' such that any agent, natural or supernatural, is able to render false a necessary truth or a true proposition about the past (*pace* Descartes). And, whatever may be the case with supernatural agents, there seems to be no sense of 'able' in which a human being is able to render a law of nature false. (But see Lewis's "Are We Free to Break the Laws?"[35] for an argument that purports to show that this is not the case – or, more exactly, for the conclusion that if a sense of 'able' satisfies all the incompatibilists' *other* requirements, then, in *that* sense human beings are able to render laws of nature false.)[36]

[34] In "A Reconsideration of an Argument against Compatibilism," *Philosophical Topics* 24 (1996): 113–122, Thomas McKay and David Johnson have shown (in effect) that there are counterexamples to the validity of this schema (they are not particularly sensitive to the sense of 'able' involved) if '*x* is unable to render *p* false' is understood as 'there is nothing *x* is able to do such that, if *x* did that thing, then *p* would be false'. These counterexamples, however, are not counterexamples to the validity of the schema if that phrase is understood as 'there is nothing *x* is able to do such that, if *x* did that thing, then *p might* be false'. The validity of the schema, understood in the latter sense, is sufficient for the apparent soundness of "the usual arguments for the incompatibility of free will and determinism."

[35] In David Lewis, *Philosophical Papers, Volume II* (Oxford University Press, 1987), pp. 291–298. The paper first appeared in *Theoria* 47 (1981): 113–121, and is available on line at www.andrewmbailey.com/dkl/Free_to_Break_the_Laws.pdf

[36] Interestingly enough, this argument had nothing to do with Lewis's Humean conception of laws. I have often heard philosophers express puzzlement that Lewis did not appeal to the Humean conception of laws in his defense of compatibilism. *I* am not puzzled. That a true universal proposition represents a "mere exceptionless regularity" hardly implies that human beings are able to render it false. Suppose, for example, that the most massive star is 260 times as massive as our sun. 'All stars have masses less than or equal to 260 solar masses' may well be a mere exceptionless regularity: it may well be that there could have been a star with a mass of 261 solar masses. But, no doubt, no human being is (or ever has been or ever will be) able to cause a counterinstance to this regularity to exist. I would suppose that, even if Lewis's Humean conception of laws is right, it would be an even more difficult task to produce a counterinstance to a *law* – in the sense in which it would be "even more difficult" for me to lift an object weighing 10,000 kilograms than it would be for me to lift an object weighing 1,000 kilograms. I am certain that these considerations are more or less those that Lewis would have adduced if he had been asked why he did not appeal to his Humean conception of laws in his defense of compatibilism.

There are obviously senses of 'able' such that if 'able' is understood in any of those senses (taking into account the point mentioned in note 34) the schema is invalid. The "skill" or "general" sense of ability and the Conditional Sense are two. Now obviously the "skill" sense of ability is not a sense that has much relevance to the problem of free will and determinism: no one would say that Grigory Sokolov now has a free choice about whether to play Chopin's Prelude in E Minor if he is now a castaway on a pianoless desert island, not even if he has the piece in his fingers, as pianists say. Nor would anyone say that it was "within his power" to play the Prelude in E Minor. There seem, moreover, to be arguments that prove decisively that free will requires more than the ability to do otherwise than one does in the Conditional Sense.[37]

Matters are otherwise when we turn to the question whether an undetermined act can be free. Suppose that our friend Alice is once more deliberating about whether to lie, to tell the truth, or to remain stubbornly silent. Suppose she is able$_{RS}$ to do each of these things. Then, if free will is incompatible with determinism, it is undetermined whether she will lie, will tell the truth, or will remain silent. It is a consequence of libertarianism that it is undermined which of these things she will do *and* that she is able$_{RS}$ to do each of them. For an examination of the difficulties that face this consequence of libertarianism, the reader is directed to Chapter 11 of this volume.

[37] It may be that in the closest possible worlds in which one chooses to *X*, one has abilities one does not have in actuality – and it may be that the ability to *X* is one of them. To take an extreme case, suppose that Alice is in a medically induced coma – but has no "long-term" motor disabilities. It may be that in all the possible worlds closest to the actual world in which she chooses to walk, she is conscious in the normal sort of way and, as an immediate consequence of her choice to walk, she walks. According to the usual understanding of counterfactual conditionals, it is true – true in the actual world – that if she chose to walk she would. But obviously she is unable to walk: she does not have a free choice about whether to walk; it is not within her power to walk. One could hardly imagine a clearer case of someone unable to walk than a person in a medically induced coma.