

FROM DETERMINISM TO RESIGNATION; AND HOW TO STOP IT¹

RICHARD HOLTON, MIT

Abstract: A few philosophers have held that determinism should lead to an attitude of resignation: since what will be will be, there is no point trying to influence the future. That argument has rightly been seen as mistaken. But a plausible parallel argument leads from the thesis of predictability—the thesis that it can be known what will happen—to an attitude of resignation. So if predictability is true, our normal practical attitudes may well be deeply mistaken. Fortunately, whilst determinism is a plausible doctrine, there is a strong argument against predictability. Determinism and predictability should be kept well apart.

DETERMINISM & PREDICTABILITY

It could well be—the science is complicated and uncertain—that the world is deterministic.² What do we mean when we say that? Something like this: that given the initial conditions, and the laws of nature, the way that the world will unroll is determined. Or, if we want to avoid using the idea of determination within the definition, we might say this: that any two situations with identical beginnings working under identical laws will unroll in the same way.³

Alternatively we might put the point in terms of knowledge. On this approach, to say that the world is deterministic is to say that if one knew the initial conditions and the laws of nature, then one could, given enough time and computing power, work out what would happen.

I shall reserve the term *determinism* for the first of these theses; the second I'll call the thesis of *predictability*. Predictability seems to require determinism: how could knowledge of the initial conditions and laws enable one to know what will happen if they did not determine what will happen? Whether there is an entailment in the other direction is less obvious. Before investing that issue though, let us identify another pair of doctrines.

FATALISM & RESIGNATION

Oedipus was fated to kill his father and marry his mother. What do we mean when we say that? Certainly it must have been true that Oedipus would kill his father and marry his mother. But at least on one understanding, the claim seems to involve more: it involves the idea that there was nothing that Oedipus could have done that would have stopped him from killing his father and marrying his mother. Somehow, no matter what he chose to do, no matter what actions he performed, circumstances would conspire to guarantee those

¹ Thanks to the members of the graduate class at MIT where I developed some of these ideas, to the audience at Edinburgh who discussed an early draft, and to Rae Langton, Damien Rochford, Stephen Yablo and the editors for useful conversations. Special thanks to Caspar Hare for much input; some day we might write a better version of this paper together.

² For a review of the science see (Earman, 2004).

³ (Lewis, 1983).

outcomes. Fatalism, understood this way, thus amounts to powerlessness to avoid a given outcome.⁴

We can put the point in terms of a counterfactual:

There are some outcomes such that whatever action Oedipus were to perform, they would come about

or more formally:

$[\exists x: \text{outcome } x] [\forall y: \text{action } y] (\text{If Oedipus were to perform } y, \text{ then } x \text{ would come about})$

This is a very specific fatalism: two specific outcomes are fated. There is no implication that Oedipus could not effectively choose to do other things: he was free to choose where he went, what he said, what basic bodily actions he performed. But we can imagine Oedipus's choices being progressively constrained, so that more and more outcomes become fated. How far could the process go? We could certainly imagine a case in which he only retained control of his basic bodily actions; all the distal outcomes that were the further consequences of those actions would be fated. That gives us:

$[\forall x: \text{actual outcome } x] [\forall y: \text{possible action } y] (\text{If Oedipus were to perform } y, \text{ then } x \text{ would come about})$

We could go further and imagine a global action fatalism, where Oedipus's choices would have no impact on his actions, even the most basic. In other words, whatever Oedipus chose, his actions would be the same:

$[\forall x: \text{actual action } x] [\forall y: \text{possible choice } y] (\text{If Oedipus were to make } y, \text{ then Oedipus would perform } x)^5$

I'll call counterfactuals like these three *Powerlessness Counterfactuals*.

Could we go further still, and imagine a form of fatalism in which Oedipus had no control even over his choices? Here things become unclear. We have made sense of Oedipus's powerlessness in terms of the lack of impact of something that is in his control on something else: his choices and actions lack the power to affect certain outcomes. Once he loses control of everything, we can't see it this way; and once he loses control of his choices, it looks as though he has lost control of everything. Perhaps there is some way of understanding a totally global fatalism, but I don't see what it is. So let us go no further than the idea of losing control over some or all of one's actions, as that is captured by the powerlessness counterfactuals. That will be far enough.

Oedipus knew his fate, but knowledge is not essential to fatalism. One can be fated without knowing one's fate, or without knowing that one is fated at all. Indeed, at least one prominent social psychologist has suggested that choice is never causally effective—that

⁴ I don't say that this is the only way to understand it; we might think that part of the poignancy of Oedipus's story is that he could have avoided his fate if only he hadn't tried to do so. To this extent then, my use of the term 'fatalism' will be stipulative.

⁵ I have here defined powerlessness in terms of the powerlessness of choices, which I find to be the most intuitive way of characterizing the idea. Alternatively it could be put in terms of the powerlessness of desires, or of judgements. This would not affect the points made here.

something like the third powerlessness counterfactual is true—even though most of us are convinced that our choices are effective.⁶

Suppose though that one comes to realize that a certain outcome is fated. Does it still make sense to strive to frustrate one's fate? I shan't go so far as to claim that one is rationally *bound* to bow to one's fate; certainly we might admire the resolve of the person who did not, and might not judge them irrational. But once one realizes that one will have no effect on the outcome, it is surely at least rationally *permissible* to stop trying to control the outcome. More broadly, even if one doesn't know *which* outcome is fated, the knowledge that *some* outcome is fated seems to give one rational permission to stop trying to control the outcome, since one knows that such activity will be pointless. Knowledge of fatalism as we are understanding it thus legitimates a fatalistic attitude in the popular sense: the view that since there is nothing that one can do to affect the future, there is no sense in trying. Without trying to define it rigorously, let us call such a position *Resignation*.

RELATING THE DOCTRINES

So we have, on the one hand, determinism and predictability; and, on the other fatalism, and the attitude of resignation. My interest is in the relations that hold between these various positions. In particular, conceding that knowledge of fatalism rationally legitimates resignation (if you know that an outcome is fated, there is no point trying to influence it), my interest is in whether whether there is a path from either determinism or predictability to resignation.

A few philosophers, typically those who are opposed to determinism, have held that belief in determinism legitimates resignation: that if you think the world is determined you are justified in not trying to influence it. Critics of the Stoics took this line in propounding the Lazy Argument, and some influential recent writers have made similar claims. More broadly there is some evidence, as we shall see, that ordinary people tend to conflate determinism and fatalism, and so tend to move directly from a belief in determinism to resignation.

Nonetheless, I think that most philosophers have held that moving from determinism to resignation is simply a mistake. Resignation is only justified if one thinks that one is powerless to affect the future—in effect, if fatalism is true—and determinism gives no grounds to think that. I agree, for reasons that I shall outline.

But that is not the end of the matter. For whilst one cannot move from determinism to resignation, I will argue that things are rather different for the move from *predictability* to resignation. That is, there is an argument that takes us from the premise that an outcome can be predicted, to the conclusion that there is no point in trying to influence that outcome, and hence that an attitude of resignation is appropriate. I am not completely sure that the argument is a good one; but it not obviously flawed. It certainly does not fall foul of the standard objections that can be raised against the move from determinism to resignation.

I suggest that this is important. Predictability is frequently run together with determinism. How many introductory classes on free will move without comment from the idea that everything that happens is determined, to the idea that if you knew enough you

⁶ (Wegner, 2002) and the work of Benjamin Libet on which Wegner draws. See (Mele, 2010) for a convincing rebuttal of this and other similar positions. For more nuanced conclusions from similar considerations see Haynes, this volume.

could predict what will happen?⁷ Certainly in the classes I have taught I have often been guilty of such a slide. And if this is true of philosophers, whose job it is to be careful about such things, isn't it likely to be true of many other people who have given it less thought? It is little wonder then that people tend to move from a belief in determinism to a belief in fatalism and to an attitude of resignation; for they may be conflating determinism with predictability.

The right response is to distinguish more clearly between determinism and predictability. In the last section I give further grounds for doing so. We should not just reject predictability because it gives a plausible route to resignation. There is a more direct argument for thinking that, in a world containing reflective beings like ourselves, predictability must be false.

MOVING FROM DETERMINISM TO RESIGNATION: SOME EXAMPLES

I start with the argument that I reject, the one that takes us from determinism to resignation. It will be useful to have some examples. The Stoics believed in determinism. In response, critics argued that they were committed to resignation:

If it is your fate to recover from this illness, you will recover, regardless of whether or not you call the doctor. Likewise, if it is your fate not to recover from this illness, you will not recover, regardless of whether or not you call the doctor. And one or the other is your fate. Therefore it is pointless to call the doctor.⁸

Since it is patently not pointless to call the doctor, the critics concluded that determinism must be false.

As an argument against determinism this is not terribly compelling, since it seems simply to equate determinism with fatalism. The critics allege that in believing in determinism the Stoics are committed to certain outcomes being fated; to embracing something like the powerlessness counterfactuals. But we need an argument for such an attribution, and the critics failed to provide one. Others have tried to do better. Recognizing the difference between the views, some more recent philosophers have tried to give an argument that determinism nonetheless entails fatalism. Richard Taylor writes:

A determinist is simply, if he is consistent, a fatalist about everything; or at least, he should be. For the essential idea that a man would be expressing by saying that his attitude was one of fatalism with respect to this or that event—his own death, for instance—is that it is not up to him whether, or when or where this event will occur, that it is not within his control. But the theory of determinism, as we have seen, once it is clearly spelled out and not hedged about with unresolved “ifs”, entails

⁷ Even when the distinction is noted, little importance is placed on it. The much used collection *Reason and Responsibility* says that determinism and predictability are ‘roughly equivalent’ whilst noting the distinction in a footnote. (Feinberg & Schafer-Landau, 1999) p. 410.

⁸ (Cicero, 1987). For a detailed examination of the argument as the Stoics faced it, see (Bobzien, 1998).

that this is true of everything that ever happens, that it is never really up to any man what he does or what he becomes, and that nothing ever can happen, except what does in fact happen.⁹

He goes on to say that fatalism should lead to resignation:

A fatalist is best thought of, quite simply, as someone who thinks he cannot do anything about the future. He thinks it is not up to him what will happen next year, tomorrow, or the very next moment. He thinks that even his own behaviour is not in the least within his power, any more than the motion of distant heavenly bodies, the events of remote history, or the political developments in faraway countries. He supposes, accordingly, that it is pointless for him to deliberate about anything, for a man deliberates only about those future things he believes to be within his power to do and forego.

The idea of resignation is here put in terms of deliberation, but presumably the same holds of choosing and of striving.

This way of thinking is not limited to philosophers. The psychologists Roy Baumeister, William Crescioni and Jessica Alquist also characterize deterministic beliefs in terms of a kind of resignation:

To the lay determinist, everything that happens is inevitable, and nothing else was possible. Thinking about what might have been is thus pointless if not downright absurd, because nothing else might have been (other than what actually happened).¹⁰

They go on to suggest that such a response is, at the very least consistent with determinism, and perhaps that it is entailed by it:

The lack of counterfactual thinking in the no-free-will condition can be considered as a straightforward response that is consistent with determinism. After all, if nothing could have happened other than what actually happened, then there are no counterfactuals.

By 'no counterfactuals' I take that they mean there are no true counterfactuals with false antecedents, and hence that counterfactual thinking is pointless.

There is reason to think that a similar move from determinism to resignation is there in popular thought. A number of studies have demonstrated that when ordinary people are brought to believe in determinism, they are less likely to behave morally. How should we explain this? Elsewhere I have suggested that it is because they fall prey to resignation: they move from the idea of determinism to the idea of fatalism, and so become convinced that there is no point in struggling against their baser urges.¹¹ Work by Eddy Nahmias and Dylan Murray provides support for this. They find that people tend to conflate

⁹ (Taylor, 1963) p. 55.

¹⁰ (Baumeister et al. 2011).

¹¹ (Holton, 2009) Ch. 8.

determinism with what they call ‘bypassing’, that is, with the view that what determines action bypasses anything that the conscious self can do. Fatalism is a form of bypassing.¹²

WHY DETERMINISM DOESN’T ENTAIL FATALISM OR SUPPORT RESIGNATION

I take it that fatalism shouldn’t be conflated with determinism; they are clearly different views. But what of the arguments we have just seen that hold that determinism entails fatalism? Those too I take to be in error. The central idea in fatalism concerns what I called *powerlessness counterfactuals*: the claim that outcomes are not dependent on agents’ actions or intentions. Returning to the example from the Lazy Argument, fatalism means that either

If I were to call the doctor I would recover, and
If I were to not call the doctor I would recover

or

If I were to call the doctor I would not recover, and
If I were to not call the doctor I not would recover

will be true. And if one of them is true, then it does indeed seem pointless to call the doctor. Calling the doctor would only be worthwhile if it made a difference to recovery; ideally if recovery were counterfactually dependent upon it, that is, if the following two counterfactuals were true:

If I were to call the doctor I would recover; and
If I were not to call the doctor I would not recover

But the truth of those counterfactuals is quite compatible with determinism. Indeed, the central idea behind determinism is that outcomes are determined causally. And, though the exact relation between counterfactuals and causation is controversial, it seems that something like that pair of counterfactuals will be true whenever there is causation.¹³

So determinism does not entail fatalism, nor does it thereby support resignation; the Lazy argument against the Stoics does not work. For parallel reasons, it seems to me that Taylor’s argument, at least as I have understood it, does not work; it simply isn’t true that if a man embraces determinism he should conclude that for all events ‘it is not up to him

¹² (Nahmias & Murray, 2011).

¹³ The most influential counterfactual account is due to David Lewis (Lewis, 1973). For subsequent work see (Collins et al. 2004). A recent development understands the counterfactuals in terms of interventions; see (Pearl, 2000) and (Woodward, 2003) for the account, and the papers collected in (Gopnik & Schulz, 2007) for evidence of its psychological applicability.

whether, or when or where this event will occur, that it is not within his control'.¹⁴ And the same can be said of the argument that Baumeister *et al.* credit to the lay determinist.¹⁵

Indeed, it might seem that the move from determinism to fatalism and so to resignation is so transparently flawed that no thoughtful person would ever make it, and that it is quite uncharitable to attribute it to ordinary people. This is a response that I have often received when I have suggested that ordinary thinking tends to run determinism and fatalism together. Putting aside the issue of whether this is a gross mistake—what can look obvious once it is carefully spelled out in a philosophical argument need not be obvious before—I will argue that there is another route to resignation, from a starting point that is often, even amongst philosophers, taken to be effectively equivalent to determinism. Perhaps, rather than making a simple error, ordinary thinking is on to something that many philosophers have missed.

FROM PREDICTABILITY TO RESIGNATION

Fatalism leads naturally to resignation. I want to argue that predictability leads to it too. This is clearest when the future is not just predictable, but predicted. But even when the predictions could be made but have not been, there is a slippery slope argument that resignation still follows. Like most slippery slope arguments, it is not completely conclusive, but it is strong enough to raise a serious worry.

Let us return to the Lazy Argument, and to whether you should call the doctor. Let us fill in a few more details to ensure that the issues are clear. Suppose that you are seriously ill. Suppose too that you live where medicine is private and that you are poor and have no insurance, so that calling the doctor would cost a great deal of what little you have. However, you have good grounds for thinking that doctors are typically effective—you know, for instance, those who call them typically do better than those who don't. So you tend to believe of other people that the following counterfactuals hold:

If they were to call the doctor they would recover; and
If they were not to call the doctor they would not recover

Putting all this together you think that whilst both calling and not calling are open options for you (you could easily imagine yourself doing either), it is rational for you to call.

But now suppose that you simply *know* you will recover. This is not conditional knowledge: it is not merely that you know that you will recover if you call the doctor, and you intend to call the doctor. You know that you will recover *simpliciter*. Perhaps you got the knowledge by divine revelation, or from some trustworthy and uncharacteristically straightforward oracle; perhaps you have some telescope that can look into the future; perhaps you were told by a returning time-traveller, who could even be yourself; or perhaps you have some machine that, given the laws and initial conditions, can compute how

¹⁴ Perhaps I have misinterpreted Taylor; perhaps he is really aiming to give something like the argument that if the past is fixed, and the past entails the future, then the future is fixed—something along the lines of the so-called 'consequence argument'. My arguments here doesn't touch on that position. But his talk of determinism being global fatalism certainly strongly suggests the interpretation I have given.

¹⁵ I give a longer response in (Holton, 2011).

certain events will turn out.¹⁶ There are real question marks over the possibility of any of these, but put such doubts aside for now. Assume simply that you have robust knowledge that you will recover, robust enough that it can provide a fixed point in your practical reasoning.

Is it rational to try to call the doctor? I suggest that it is not; in this sense, robust predictability is like fatalism. For if you know that you will recover, then there appear to be two possibilities for you to consider:

I call the doctor and I recover; and
I do not call the doctor and I recover

Of these you clearly prefer the latter, since calling the doctor is expensive. In each case the second conjunct you know to be true. But since it appears to be up to you which of the first conjuncts you make true, and hence which of the two conjunctions is true, then it is rational to make true the one that you prefer. So it is rational not to call the doctor.

Now suppose that you robustly know in a similarly robust fashion that you will not recover. In this case the possibilities are:

I call the doctor and I do not recover; and
I do not call the doctor and I do not recover

Similar reasoning applies. You prefer the latter to the former, since at least you keep your money to make your last days comfortable and to pass on to your loved ones; so, given that you know you will not recover, it is rational not to call the doctor.

Note the difference between this case and the case in which determinism is true. Determinism entails only that you know that you will either recover or not:

Know (p or not-p)

Prediction distributes that knowledge:

Know (p) or Know (not-p)

The interesting result is that if you *either* robustly know that you will recover, *or* you robustly know that will not, it is rational to not call the doctor.

Suppose though that you do not know what will happen, but you know that it is *predictable*. This is where we encounter the slippery slope. Suppose first that whilst you do not yet know whether or not you will recover, you know that your friend knows whether or not you will; and, moreover, you know that very soon your friend will tell you. Suppose however that you have to make your decision about whether to call the doctor before you are told. Should you call? It seems to me that you might reason as follows: If I could only delay my decision until after I was told, I know whatever I was told I would decide to not

¹⁶ Does this mean that you will deny that one of the previous counterfactuals apply to you, i.e. that if you were not to call the doctor you would not recover? Well you might deny it; but equally you might not know quite what to think. For further discussion see below.

call the doctor. But since that would be a rational thing to do once I was told, that is good enough grounds for deciding not to call before I am told.¹⁷

Now for the next step along the slippery slope: it's not my friend who knows: it is rather that the prediction has been made by a machine that can predict the future, and I will not have a chance to read the prediction before I have to decide whether or not to call. The practical conclusion is as before.

Now the next step: the machine is still doing the calculations; it will not finish till after I have to decide whether to call. Same practical conclusion.

The next step: the machine has all the evidence it needs, but it hasn't yet started doing the calculations. Same practical conclusion.

Next: the machine isn't actually going to do the calculations, but it could. Same practical conclusion.

Next: the machine hasn't actually been built yet, but we know how to build it, and if it were built it could do the calculations. Same practical conclusion.

Next: we don't know how to build the machine, but we know that in principle a machine could be built that could do the calculations. Same practical conclusion.

There are many places along this slope where someone might want to call a halt. Perhaps it should be halted right at the beginning: there is all the difference in the world between knowing something yourself, and having someone else know it. One might make a case for that, but it does seem rather odd. Moreover, I think that there is a principled reason for following the slope all the way to the end. The conclusion of the slope is to move the condition that is needed for Resignation from

Know (p) or Know (not-p)

to the weaker

Knowable (p) or Knowable (not-p)

The reason that knowledge justified Resignation was that it provided a fixed point around which deliberation could move. But the same is true if an outcome is merely knowable in advance. If it is knowable in advance, then one could come to know it. And so there is a fixed point, even if one doesn't know what it is.

Thought of in this way though, we see why the slippery slope doesn't take us all the way to determinism. For determinism doesn't by itself provide us with the possibility of a fixed point that could be used in deliberation about what will happen. For determinism by itself doesn't guarantee that what will happen is knowable. One way in which it could fail to be knowable, quite compatible with determinism, is if the very process of coming to believe something would have an impact on what would happen. And it is this feature that I will use in arguing against the possibility of predictability. Before that though, let us address the question of whether predictability entails fatalism.

¹⁷ Debates on the bearing of God's knowledge of future contingents for our freedom might be expected to have some relevance here. For comprehensive discussion see (Zagzebski, 1991). In fact though I think that they are not of very much use, for a number of reasons: (i) most of the discussion is of the broader issue of the compatibility of God's foreknowledge with freedom, and not on the rationality of Resignation; (ii) God is normally taken to be both essentially knowable and immutable; (iii) there is no discussion of what difference it would make if God communicated his knowledge with human actors.

DOES PREDICTABILITY ENTAIL FATALISM?

I argued that knowledge that you will recover presents you with two possible outcomes

I call the doctor and I recover
I do not call the doctor and I recover

of which it is rational to choose the second. Now it might seem that your knowledge that you will recover, combined with your control over whether or not you call, gives you confidence in both of the following counterfactuals:

If I were to call the doctor I would recover
If I were to not call the doctor I would recover

But (assuming that there is nothing else that you can do to affect your recovery) that is just to claim that your recovery is fated. So should we conclude that predictability entails fatalism?

I think that we should not. These examples don't show it so clearly. But imagine another that does. Suppose that a coin is hidden under one of the two cups in front of you. You have to choose the cup where it is. Suppose you come to know that you will choose correctly. So there seem to be two outcomes:

I choose the left-hand cup and the coin is under the left-hand cup
I choose the right-hand cup and the coin is under the right-hand cup

Should you now go on to endorse the counterfactuals:

If I were to choose the left-hand cup, the coin would be under the left-hand cup
If I were to choose the right-hand cup, the coin would be under the right-hand cup?

Surely not. The position of the coin is fixed; my choice isn't going to move it around. Foreknowledge of the outcome does not commit one to the truth of the counterfactuals. Likewise with the doctor. It could be that recovering or not recovering is fated, so that nothing I could do would have any impact. But that is not a conclusion that I can draw from the simple knowledge of what will happen.

DENYING PREDICTABILITY

As I have said, philosophers tend to move between determinism and predictability without thinking much is at stake. I claim that, even if determinism is true, we have excellent grounds for thinking that predictability is not. And this doesn't result from anything especially sophisticated about human beings and their essential unpredictability. I think that Predictability would be equally well refuted by the existence of the kind of machine that many of my colleagues at MIT could construct in their lunch hour. Let me explain.

THE CHALLENGE

Suppose that I offer you a challenge. Your job is to predict whether the light bulb on a certain machine will be on or off at a given time. Moreover, I will give you considerable

resources to make your prediction: you can have full information about the workings of the machine, together with as much computing power as you wish. Lest that seem to make things too easy, I point out that makers of the machine had one goal in mind: to render your prediction wrong.

Of course, a natural way for the makers to construct the machine would be to equip it with a mechanism that could detect what you would predict, and then a device to ensure that the prediction was false. Still, there are things that you might do to avoid this. One technique would involve keeping your prediction a secret from the machine; another would involve making it so late that the machine would not have time to act on it. But unfortunately for you, I rule both of these options out. Your prediction has to be made in full view of the machine. I give you a light bulb of your own. You must switch it on to indicate that you predict the that the light on the machine will be on at the set time; you must switch it off to indicate that you predict the light on the machine will be off. And I require that the prediction be made a full minute before the set time. So, for instance, if you predict that the light on the machine will be on at noon, you have to have your light switched on at 11.59.

Now you seem to be in trouble. For, as we said at the outset, making a machine capable of defeating you in the challenge is surely within the capabilities of many of my colleagues, indeed many of the undergraduates, at MIT. All it would need is a simple light sensor, and a circuit that gives control of its light to the output of the sensor. If it senses that your light is on, it switches its light off, and vice versa.

At this point you might think that I have rigged the challenge against you. So let me make one final offer. I will give you, not simply all the knowledge about the workings of the machine, but all the knowledge about the workings of the universe. Moreover—don't ask how I do this—I guarantee that the universe will be determinate. And, as before, I give you as much computing power as you wish.

Should you now accept the challenge? I think not. For how can this new knowledge help? At best it seems that it will enable you to predict how you will lose: you will perhaps be able to calculate what move you will make, and how the machine will respond to it. But should you try to use that knowledge by changing your prediction to beat the machine, either you will fail to make the change, or it will turn out that what you thought was knowledge was not knowledge after all. The machine will not be concerned with your machinations. Whatever it is that you predict, it will simply control its light to ensure that your prediction is false.

I say that the best you will be able to do, despite your knowledge and your computing power, is to predict how you will lose. I actually doubt that you will even be able to do that reliably. But to see this, let me fill in some background.

My story might remind some readers of an article written by Michael Scriven, and refuted by David Lewis and Jane Richardson.¹⁸ Scriven argued that human beings are unpredictable because they have the capacity to act in a way that runs contrary to any prediction, and that was the source of the refutation. But I am not making a parallel claim about the machine that I have challenged you to predict. It would be quite possible for *someone else* to predict the machine's behavior. Indeed, it would be quite possible for you to predict the machine's behaviour, provided that you didn't reveal the prediction to the machine. What I claim that you will not be able to do is to make an accurate prediction that you reveal to the machine ahead of time.

¹⁸ (Scriven, 1964); (Lewis & Richardson, 1966). For a very useful review of various authors who argued along similar lines to Scrivens, and some helpful suggestions of their own, see (Rummens & Cuyppers, 2010). I wish that I had seen their paper before writing this one.

However, like Scriven, I do want the conclusions to apply to human beings. Obviously if they are counter-suggestible they should be just as hard to predict as the machine. But there is something more complicated about human beings, since they are able to make predictions about themselves. This means that they can play the role both of predictor and of frustrater. Suppose that Tiresias wants to predict what he will do, but also wants to ensure that he will not act in accord with his prediction. Provided that he makes his prediction clearly to himself, and well enough in advance, it will be his frustrating side, and not his predicting side, that will triumph.

This is why I doubt that you will even be able to make an accurate prediction about how your game with the machine will go. Here your motivation is not to frustrate yourself; it is to trick the machine. But that is good enough to render any calculation you make highly unreliable.

CONCLUSION

Determinism and predictability are often run together, but they should be kept well apart. Whilst determinism does not lead to resignation, predictability plausibly does. Fortunately, whilst determinism may well be true, there is a good argument against predictability, and so, even if it does entail resignation, this poses no threat to our ordinary way of thinking about things. The tangle of doctrines surrounding issues of determinism and freedom is dense. It should be no surprise if ordinary people get caught in it; philosophers are still some way from disentangling them all.

REFERENCES

- Baumeister, Roy, A. William Crescioni and Jessica Alquist, 2011, 'Free Will as Advanced Action Control for Human Social Life and Culture' *Neuroethics* 4, 1–11.
- Bobzien, Susanne, 1998, *Determinism and Freedom in Stoic Philosophy* (Oxford: Clarendon Press).
- Cicero, 1987, *On Fate* 28, in A. Long and D. Sedley (eds.) *The Hellenistic Philosophers* (Cambridge: Cambridge University Press) p.339.
- Collins, John, Edward Hall, and Laurie Paul, (eds.), 2004, *Causation and Counterfactuals* (Cambridge, Mass: MIT Press).
- Earman, John, 2004, 'Determinism: What We Have Learned, and What We Still Don't Know' in J. Campbell, M. O'Rourke and D. Shier (eds.) *Freedom and Determinism* (Cambridge MA: MIT Press) pp. 21–46.
- Feinberg, Joel and Russ Schafer-Landau, 1999, *Reason and Responsibility* 10th Edition (Belmont, CA: Wadsworth).
- Gopnik, Alison and Laura Schulz (eds.), 2007, *Causal Learning* (New York: Oxford University Press).
- Haynes, 2012, this volume.

- Holton, Richard, 2009, *Willing, Wanting, Waiting* (Oxford: Clarendon Press).
- 2011 “Comments on ‘Free Will as Advanced Action Control for Human Social Life and Culture’ by Roy F. Baumeister, A. William Crescioni and Jessica L. Alquist” *Neuroethics* 4, 13–16.
- Lewis, David, 1973 ‘Causation’, *Journal of Philosophy*, 70, 556–67, reprinted with post script in his *Philosophical Papers* Vol. II (New York: Oxford University Press, 1986) 159–213.
- 1983, ‘New Work for a Theory of Universals’ *Australasian Journal of Philosophy* 61 343–77, reprinted in his *Papers in Metaphysics and Epistemology* (Cambridge: Cambridge University Press, 1999) 8–55.
- and Jane Richardson, 1966, ‘Scriven on Human Unpredictability’, *Philosophical Studies*, 17, 69–74.
- Mele, Al, 2010, *Effective Intentions* (Oxford: Oxford University Press).
- Nahmias, Eddy, and Dylan Murray, 2011 ‘Experimental Philosophy on Free Will: An Error Theory for Incompatibilist Intuitions’ in Jesús Aguilar, Andrei Buckareff and Keith Frankish (eds.) *New Waves in Philosophy of Action* (Basingstoke: Palgrave Macmillan) pp. 189–216.
- Pearl, Judea, 2000, *Causality* (Cambridge: Cambridge University Press).
- Rummens, Stefan and Stefaan Cuypers, 2010, ‘Determinism and the Paradox of Predictability’, *Erkenntnis* 72, 233–249.
- Scriven, Michael, 1964, ‘An Essential Unpredictability in Human Behavior’, in *Scientific Psychology: Principles and Approaches*, ed. Benjamin B. Wolman and Ernest Nagel (New York: Basic Books).
- Taylor, Richard, 1963, *Metaphysics* (Englewood Cliffs NJ: Prentice Hall).
- Wegner, Daniel, 2002, *The Illusion of Conscious Will* (Cambridge MA: Harvard University Press).
- Woodward, James, 2003, *Making Things Happen* (New York: Oxford University Press).
- Zagzebski, Linda, 1991, *The Dilemma of Freedom and Foreknowledge* (New York: Oxford University Press).