

Caspar Hare
Richard Holton

DRAFT

April 2012

Determinism, Sloth and the Opacity of the Future

1.1 The Lazies

There is some evidence¹, formal and informal, that believing that determinism is true demotivates people to work to bring about outcomes that they take to be desirable. Furthermore the move from believing determinism to demotivation appears to be mediated by something that philosophers would recognize as *reasoning*.

What might be going on? It might be that the reasoning has to do with moral responsibility: In the thrall of a protestant work ethic, I work only to avoid the guilt, sin and general blameworthiness that attach to sloth. I come to believe that determinism is true. Being, at heart, an incompatibilist about determinism and moral responsibility, I infer that no guilt, sin and general blameworthiness attach to sloth. So I take myself to have no reason to work.

But that will not explain why people are demotivated to work in contexts in which they clearly have some guilt-independent incentive to work.

One hypothesis is that, in these contexts, people are gripped by what philosophers call the *Lazy Argument*.² Suppose their situation is something like my situation here:

The Big Exam

I don't enjoy studying for exams. All other things being equal, lazing beats working, in my book. But a lot rides on the result of this exam, so my preferences are like this:

I laze, and pass
I work, and pass

¹ Summarized in chapter 8 of Richard Holton, *Willing, Wanting Waiting* (Oxford: Clarendon Press, 2009).

² The argument was first made by critics of Stoic determinism. For a detailed examination of the criticism see Suzanne Bobzien, *Determinism and Freedom in Stoic Philosophy* (Oxford: Clarendon Press 1998).

BIG GAP

I laze, and fail

I work, and fail

And I think that I will be much more likely to pass if I work.

Ordinarily I would be moved to work, but now I take seriously the idea that determinism might be true, and I think to myself...

Powerless

If determinism is true, then I my study-habits will not affect the result of the exam.

Dominance

If my study-habits will not affect the result of the exam then it makes sense for me to laze – because, for any way that this result that I will not affect might be, I would rather that I laze and things be that way, than that I work and things be that way.

Lazy

If determinism is true then it makes sense for me to laze.

There is an obvious problem with this argument. *Powerless* is false.³ It does not follow from the truth of determinism that my study-habits will not affect the result of the exam. Maybe I will work and thereby cause myself to pass. Maybe I will laze and thereby cause myself to fail. Maybe I will work and (in some odd way) thereby cause myself to fail. Maybe I will laze and (in a similarly odd way) thereby cause myself to pass. There can be relations of cause and effect between events in a deterministic world.

So it looks as if the great unwashed are making a simple blunder. A few carefully chosen remarks about causation will set them straight.

It seems to us that this is unfair to the great unwashed. There is a much more interesting argument to the same conclusion. The conclusion remains false, you will

³ There is an unobvious problem with the argument too – evidential decision theorists (see section 1.4) will reject *Dominance*.

doubtless be relieved to hear. But we find the argument much more interesting because, to us, it is not at all obvious where it goes wrong. We think that a close look at why and where it goes wrong reveals some curious things about the scope of reflection principles in the theory of practical rationality, and about our ability to know our own future behavior.

1.2 The New Lazies

What is Determinism? The claim concerns a relation between the precise way the world is at some times and the precise way the world is at other times. Roughly:

Determinism

The precise way the world is at any given earlier time determines the precise way the world is at any given later time.

What is the ‘precise way the world is’ at a time? It includes all temporary intrinsic features of the world at the time, all laws of nature that govern the world at the time, and nothing else. What is it for one way for the world to be at a time to ‘determine’ another? In English the term ‘determine’ has a modal flavor (‘the vote was a fix, the election committee determined the result in advance’), and an epistemic flavor (‘Poirot determined that the estranged husband did it’). So there are two natural ways to spell it out:

Nomological Determinism – A Claim About Nomological Possibility

The precise way the world is at any given earlier time leaves only one possible (in the sense of: *consistent with the laws*) way for it to be at any given later time.

Epistemic Determinism – A Claim About What is Knowable

It is in principle possible to know the precise way the world will be at any given later time, by knowing the precise way the world was at any given earlier time.

In public philosophers tend to chop and change between the two spellings-out – harmlessly enough, in their view, because the first sort of determinism entails the second sort of determinism.⁴ But the second sort of spelling-out has a shock-value that the first does not, so it tends to get more attention, and need more explaining. A representative explaining:

Philosopher: It is in principle possible to know precisely how the world will be at any given later time, by knowing precisely how the world was at any given earlier time.

Layperson: “Does this mean that I can know the future? Can I know the result of next year’s Kentucky Derby?”

Philosopher: “Well, let’s not get too excited here. I am not saying that it is practically possible, with present or future technology, for you to know all aspects of the future. I am talking about what is possible in principle. If you somehow came to know the precise state of the universe at an earlier time, and you somehow gained a rich enough grasp of the laws of nature that you (perhaps with the aid of some fantastic machine) could work out how the universe must evolve from that time, then yes, you could know the result of next year’s Kentucky Derby.”

As a layperson, thinking about my upcoming exam and what the philosopher just taught me about Determinism, I might be forgiven for inferring:

Foreknowledge of the Result is Attainable

If Determinism is true then it is in principle possible for me, by learning enough about past states of the world, and about the laws that govern the world, to come to know the result of the exam now, before the exam has taken place, before I have decided whether to work or laze.

⁴ This view goes back at least as far Laplace: “We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.” *A Philosophical Essay on Probabilities* 1816.

What would happen if I came to know this thing that it is in principle possible for me to know? What would it make sense for me to do then? Well, there are just two things that I might come to know – that I will fail, or that I will pass. It seems obvious that

Foreknowledge of Failure Renders Work Senseless

If I were to come to know now that I was going to fail the exam, then it would not make sense to work.

Why work if I am certain I will fail, and I regard working-and-failing as the very worst thing that can happen? And it seems obvious that

Foreknowledge of Success Renders Work Senseless

If I were to come to know now that I was going to pass the exam, then it would not make sense to work.

Why work if I am certain I will pass, and I regard working-and-passing as worse than lazing-and-passing? So it seems obvious that

Foreknowledge of the Result Renders Work Senseless

If I were to come to know now the result of the exam, then it would not make sense for me to work.

And this looks like a robust principle concerning foreknowledge:

Result Reflection

If I know that it is in principle possible for me to come to know the result of the exam now, and I know that if I knew the result of the exam now then it would not make sense to work, then it does not make sense to work.

The dread conclusion follows:

Lazy

If I know that determinism is true, then it does not make sense to work.

1.4 Foreknowledge Redux

Something must have gone wrong somewhere? Where? Perhaps the mistake came with *Foreknowledge of the Result Renders Work Senseless*. We said this was ‘obvious’ but perhaps that was hasty.

The typical way for me to learn the result of an exam before deciding whether to study for it involves my learning that my study-habits will have no causal influence over the result of the exam – “It’s a fix! The professor hates you. He is going to fail you whatever you do.” or “No worries! You have done enough work already. You will pass, no matter whether you work or laze this evening.” In such cases it may be obvious that it makes sense for me to laze. But what if I just came to learn the result of the exam, in a way that leaves it open whether my study habits have a causal influence over the result of the exam?

The case now has a curious feature. There are things inside my control – including whether I will work or laze. And there are things outside of my control – including whether if I were to work then I would pass, and whether, if I were to laze then I would fail. And my confidence that the things outside of my control are one way or another depends on my confidence that the things inside my control are one way or another.

Suppose, for example, that I learn that I will fail the exam. If I believe that I will laze (something inside my control, notice), then my confidence that if I were to work then I would fail (something outside my control, notice) remains high. But if I believe that will work (something inside my control) then my confidence that if I were to work then I would pass (something outside my control) goes down.

Now, philosophers have devised two different decision theories, *causalism* and *evidentialism*, that come apart in cases in which my confidence that things outside my control are one way or another depends on my confidence that things inside my control are one way or another.

Very roughly, causalists recommend that I act so as to make it the case that, supposing things outside of my control are the way I believe them to be, it will turn that I did the best I could. Very roughly, evidentialists recommend that I act so as to maximize my confidence that good things will happen.

Let's try to be more precise. As a background, let's suppose that your present cognitive (which is to say *belief-like*) attitudes can be represented by a function, C , from propositions to real numbers between 0 and 1 – the numbers representing how likely you think it that the propositions are true. And let's suppose that your present conative (which is to say *desire-like*) attitudes can be represented by a function, U , from propositions to real numbers – the numbers representing how desirable you think it that the propositions be true. And let's suppose that the ways in which you might act now can be represented by a set of propositions A . Call the propositions in A *act-propositions*.

Now let D be a set of propositions concerning how things beyond your control are. Let D be *exclusive* (no two propositions in D can both be true), *exhaustive* (all propositions about how things beyond your control are entail the disjunction of the propositions in D) and *relevant* (for all act-propositions a , all propositions d in D , and all propositions r , if $a \wedge d$ is consistent with both r and $\neg r$ then $U(a \wedge d \wedge r) = U(a \wedge d \wedge \neg r)$.) Call the propositions in D *dependency hypotheses*.

Where d is a variable ranging over dependency hypotheses, we define the *causal expected utility* ($E_C U$) of an act proposition, a , like this:

$$E_C U(a) = \sum_d (C(d) \cdot U(a \wedge d))$$

And we define the *evidential expected utility* ($E_E U$) of an act proposition, a , like this:

$$E_E U(a) = \sum_d (C(d/a) \cdot U(a \wedge d))$$

Where ' $C(d/a)$ ' refers to your conditional credence in dependency hypothesis d , given that you make true act-proposition a .

Causalists say that you subjectively ought to make true the act-proposition with highest causal expected utility. Evidentialists say that you subjectively ought to make true the act-proposition with highest evidential expected utility. We can get a sense of how the theories work by applying them to foreknowledge cases.

The evidentialist treatment of foreknowledge cases is straight-forward. Before I learn the result of the exam, my conditional credence that I will pass if I work is sufficiently high, and my conditional credence that I will pass if I laze sufficiently low, that the expected evidential utility of working is higher than the expected evidential utility of lazing. When I come to know that I will fail, the evidential expected utility of my working is the utility of my working and failing (because my conditional credence that I will fail if I work goes to 1), while the evidential expected utility of my lazing is the utility of my lazing and failing (because my conditional credence that I will fail if I laze goes to 1). The latter is higher, so I ought to laze. And when I come to know that I will pass, the evidential expected utility of my working is the utility of my working and passing, the evidential expected utility of my lazing is the utility of my lazing and passing. The latter is higher, so I ought to laze. Either way I ought to laze.

The causalist treatment of the cases is much less straight-forward.⁵ There are four relevant dependency hypotheses:

- D1 If I were to work then I would pass. If I were to laze then I would pass.
- D2 If I were to work then I would pass. If I were to laze then I would fail.
- D3 If I were to work then I would fail. If I were to laze then I would pass.
- D4 If I were to work then I would fail. If I were to laze then I would fail.

Before learning the result of the exam, my confidence in D3 is sufficiently low, my confidence in D2 sufficiently high, that working has higher causal expected utility than lazing.

When I come to know that I will fail things change. My credence in D1 goes to nothing. My credence in D3 stays very low. What happens to my credence in D2 and D4? That depends on what I believe I will do. If I believe I will work then my credence in D2 goes to nothing, and lazing has higher causal expected utility than working. If I believe I will laze then I retain substantial credence in D2, and working has higher causal expected utility than lazing.

So, when I come to know that I will fail, causal decision theory has it that I face what we call *self-frustrating decision dependence*⁶. If I believe I will work then I ought to laze. If I believe I will laze then I ought to work.

When I come to know that I will pass things change in a different way. My credence in D4 goes to nothing. My credence in D3 stays very low. What happens to my credence in D1 and D2? That depends on what I believe I will do. If I believe I will work then I retain substantial credence in D2, and working has higher causal expected utility

⁵ Indeed we should mention that both of us reject the causalist treatment of these cases. In these cases, causalism says that what I ought do depends on what I believe I will do. But what I ought to do never so-depends on what I believe I will do.

⁶ For details, see Hare and Hedden, 'Self-Reinforcing and Self-Frustrating Decisions' ms.

than lazing. If I believe I will laze then my credence in D2 goes to nothing, and lazing has higher causal expected utility than working.

So, when I come to know that I will pass, causal decision theory has it that I face what we call *self-reinforcing* decision dependence. If I believe I will work then I ought to work. If I believe I will laze then I ought to laze.

Either way the following is true:

Intending to Laze is No Less Rationally Defensible than Intending to Work
If I form a settled intention to laze then I will be no more or less an appropriate object of rational criticism than if I form a settled intention to work.

When I come to know that I will fail, then I will be an appropriate object of rational criticism if I form a settled intention to laze ('You intend to laze, so you believe you will laze, so you ought not to laze, so you intend to do something you ought not to do – boo!). And I will be an appropriate object of rational criticism if I form a settled intention to work ('You intend to work, so you believe you will work, so you ought not to work, so you intend to do something you ought not to do – boo!) When I come to know that I will pass, then I will not be an appropriate object of rational criticism if I form a settled intention to laze ('You intend to laze, so you believe you will laze, so you ought to laze, so you intend to do something you ought to do – good job!) And I will not be an appropriate object of rational criticism if I form a settled intention to work ('You intend to work, so you believe you will work, so you ought to work, so you intend to do something you ought to do – good job!)

What attitude is it appropriate for me to take towards two courses of action when I know that, if I form a settled intention to take the one, then I will be no more or less an

appropriate object of rational criticism than if I form a settled intention to take the other? This is a very subtle question. We won't try to answer it here. Call the attitude, whatever it is, attitude A. We can now run the New Lazy argument for causalists by appealing to

*Result Reflection**

If I know that it is in principle possible for me to know the result of the exam now, and I know that if I knew the result of the exam now then it would make sense for me to take attitude A towards working the question of whether I will work or laze, then it makes sense for me to take attitude A towards the question of whether I will work or laze.

This will demotivate me to work to the extent that attitude A is an attitude of 'anything goes'.

1.4 Reflection Redux

Perhaps, then, the problem is with the reflection principle. *Result Reflection* is an instance of a more general principle:

Practical Reflection

If I know that it is in principle possible for me to come to know something now, and I know that if I knew that something then it would make sense for me to ϕ , then it makes sense for me to ϕ .

This principle is meant to capture the attractive idea that it is rational, other things being equal, to defer to my epistemic superiors. But there are some problems with it, as stated.

One has to do with predictable preference shifts. What if I know that coming to know something would change my underlying preferences (in the way that coming to know enough about a psychopath's background might make me sympathize with him)? We need to stipulate that what matters is what it would make sense for me to do if I knew more, but retained my present underlying preferences.

Another has to do with the way in which my knowing more might change the consequences of my behaving in one way or another. Suppose I knew the answers to the exam. Then it would make sense for me to laze. But it does not follow that it makes sense for me now, not knowing the results of the exam, to laze. We need to stipulate that what matters is not what it would make sense for my epistemically superior self to do, but what it would make sense for my epistemically superior self to want my epistemically inferior self to do.⁷

Another has to do with the fact that principle picks up, not just on what I would want if I were better-informed and *fully-informed*, but also on what I would want if I were better-informed and *partially informed*. Partial information can be misleading. Your coin has been tossed many times. I don't know how it has landed each of these times. But I do know that if I were to know of all and only the first hundred head-tosses (if it happened that the first, third, sixth, seventh... one hundred and ninety fifth tosses came up heads, and it happened that I asked about only those tosses) then I would quite reasonably take the view that that coin is biased and want myself to bet, at very short odds, on its landing heads next time. It follows from the principle, as stated, that it makes sense for me to bet, at very short odds, on its landing heads next time. And it also follows from the principle, as stated, that it makes sense for me not to bet, at very short odds, on its landing heads next (because, if I knew of all and only the first hundred tail-tosses, in just the same way, I would reasonably want myself to bet against its landing heads next time.) But that is crazy.

⁷ Following Michael Smith (*The Moral Problem* 1994 Chapter 5) this has been known as an ideal *advisor* theory, contrasted with the ideal *exemplar* theory. The terminology is a bit misleading because, as Smith acknowledges, the issue is not what it would make sense for an epistemically superior twin of me to advise me to do if we somehow occupied the same world. The issue is what, if I were epistemically superior, it would make sense for me to want myself to do if I were epistemically inferior.

How do we solve the misleading-information-problem? One solution is to restrict *Practical Reflection* so that it covers only fully informed advisors – I should defer to the omniscient. This weakens the principle considerably. Another, less drastic solution makes use of an idea from epistemology: We cannot control what it will be rational for us to believe by deliberately acquiring misleading information. If I know now that, once I go down a certain evidence-gathering path, then I will have a certain attitude if I am rational throughout, then I cannot dismiss that attitude as based on misleading information. What we need to do is to restrict the principle so that it covers only better-informed advisors whose positions I can now deliberately put myself in, by going down particular evidence-gathering paths.

Let's make this precise. Say that a doxastic state is *accessible* to me when there is a way for me to gather evidence, such that I know that if I choose to gather evidence in that way, and I remain epistemically rational throughout, then I will end up in that state. What matters is what I know it would make sense for me to want if I were in accessible doxastic states. In the coin toss case, before I have asked about any of the past tosses, the state of taking the coin to be head-biased is not accessible to me, because I do not know that, by asking questions in any particular way and remaining epistemically rational throughout I will end up in that state (if I start asking questions about particular coin tosses then maybe, by fluke, I will ask about all and only the head-tosses – but I can't guarantee that.)

Altogether, this gives us have:

Practical Reflection

If there is a doxastic state D such that:

- (i) D is epistemically superior to my present state (it involves my knowing everything I know now, and more.)

- (ii) D is accessible to me-now (I know, for some way of gathering evidence, that if I gather evidence that way and remain rational then I will be in D).
 - (iii) I know that if I were in D, and I retained my present underlying preferences, then it would make sense for me to want my epistemically inferior self to ϕ .
- then it makes sense for me to ϕ .

Does this refined principle give the New Lazies what they want? The question is whether, if determinism is true, the doxastic state of knowing the result of the exam is accessible to me now, before I have decided whether to work. And the answer would seem to be yes. As the philosopher said, if determinism is true then I know, for a particular way of gathering evidence (learning enough about the initial conditions of my world and the laws of nature that govern it) that if I gather evidence that way then I will put myself in a situation in which I know the result of the exam.

Of course, as the philosopher also said, gathering evidence in that way is, practically speaking, beyond me. But that does not render the doxastic state inaccessible for the purposes of reflection. We have no principled reason to distinguish these sorts of cases:

The Paper

An oracle wrote the result of the exam on a piece of paper and placed it face down in front of me. I can very easily turn the paper over and find out whether I will pass.

The Anvil

An oracle etched the result of the exam on an iron anvil and placed the anvil face down before me. In practice the anvil is too heavy for me to lift. But if I had access to heavy machinery (which I don't) then I could find turn the anvil over and find out whether I will pass.

The Deterministic World

No oracle wrote down the result of the exam. But, because determinism is true, if I had access to heavy computational machinery (which I don't) then I could find out whether I will pass.

In all these cases there is way to gather evidence such that I know that if I gather evidence that way and remain rational then I will end up in situation in which it is rational for me to laze. In all these cases, if it is rational for me, prior to gathering the evidence, to work, then it is rational for me to protect myself against my own, better informed, rational choices, by avoiding gathering the evidence, even *paying* to avoid gathering the evidence – after all, I know that if I gather the evidence then I will laze, and I don't want to laze, I want to work. But the thought behind *Practical Reflection* is that, while there may be cases in which it is rational to avoid gathering evidence (say I know the Mafia will shoot me if I look into things too closely), there are never cases in which it is rational to protect myself against my own, better informed, rational choices. So it cannot be that it is rational for me, prior to gathering the evidence, to work.

Result Reflection and Foreknowledge of the Result Renders Work Senseless appear to hold up. The remaining premise of the New Lazy argument is the first: *Foreknowledge of the Result is Attainable*. We say that premise is false. To explain why we will look back to an old philosophical problem concerning predictability.

PART II

2.1 Predictability

The following three claims are inconsistent. It cannot be that they are all true.

- (1) The world contains a *reporter*, whose job it is to report on whether a particular event, *e*, occurs at a particular time, *t*. The reporter does not make a mistake. If it says 'Yes' then *e* occurs at *t*. If it says 'No' then *e* does not occur at *t*.
- (2) The reporter says something before *t*. It makes a prediction.

- (3) The world contains a *frustrator*, whose job it is to ensure that any predictions made by the reporter are inaccurate. The frustrator does not make a mistake. If the reporter says ‘Yes’ before t then e does not occur at t . If the reporter says ‘No’ before t then e occurs at t .

But we can easily be induced to overlook the inconsistency. Sometimes, when we take some of them to be true, and then try to move a step further, try to take them all to be true, we encounter little of the resistance, little of the feeling of mental-metal grinding against mental-metal, that typically accompanies efforts to accept the impossible.

This observation was the inspiration for a large philosophical literature in the mid-twentieth century⁸. It was thought that, by reflecting upon it, we could discover deep and surprising things about deliberation, and about what it is possible for us to know.

2.2 A Surprising Thing About the Phenomenology of Deliberation

If (1) and (2) are true, then (3) must be false. But if you yourself are in a position to be a frustrator, then sometimes, after taking (1) and (2) to be true, you may find it natural to take (3) to be a live *practical* possibility. Here is an adaptation of a story by Alvin Goodman⁹:

Alvin Goldman and the Book of Life

At a certain point in Goldman’s middle-age he comes upon a book -- *The Book of Goldman’s Life*. For every minute of his life it contains an accurate description of what he does that minute. He reads through the early sections with great care and joy, delighted to discover so much about his past. After a while the book becomes a little tedious – minute after minute of “Goldman reads this book.” Undeterred, Goldman reads ahead, and finds... predictions.

⁸ The literature started with Karl Popper, ‘Indeterminism in Quantum Physics and in Classical Physics’ *The British Journal for the Philosophy of Science* vol. 25, no. 4 (1950). For a survey of its course see Stefan Rummens, Stefaan Cuypers ‘Determinism and the Paradox of Predictability’ *Erkenntnis* vol. 72, no. 2 (2010).

⁹ From Alvin Goldman ‘Actions, Predictions and Books of Life’, *American Philosophical Quarterly*, vol. 5, no. 3 (1968).

Some are happy. Some are sad. The last is intolerable. Appalled, defiant, he resolves to frustrate the predictions. But somehow he always fails.

Sometimes he forgets the prediction (e.g. “Goldman forgets this prediction”). Sometimes he confuses predictions he has read with intentions he has formed (e.g. “Goldman orders himself a blueberry ice cream”). Sometimes he deems it too costly to defy the prediction (e.g. “Goldman dodges the car”).

Sometimes the prediction reminds him of something important that he would otherwise have forgotten (e.g. “Goldman meets with the President”). Finally he becomes weary of all the evasion, and he meekly acquiesces in his predicted suicide.

We may suppose that, fairly soon after getting to the predictive bit of the book, Goldman becomes justifiably convinced that it is indeed a book of his life, that every sentence in the book is accurate. But, right up to his suicide, this conviction does not prevent him from deliberating about whether to frustrate the predictions. He knows that the book predicts he will meet with the President. He knows that the book is accurate. Yet he still thinks carefully about not meeting with the President, and he decides that, on balance, it is a bad idea.

The lesson that Goldman wanted us to draw from stories like this was that foreknowledge of what you will do does not block deliberation about what you will do. An option, like not-meeting-with-President, can be a dead epistemic possibility for you, though it remains a live practical possibility for you.¹⁰

¹⁰ The same point can be made with the grandfather paradox for one-world-history time travel. In a one-world-history time travel scenario it cannot be that you travel to the past, aim a gun at your grandfather’s head, pull the trigger, and kill him, thereby preventing him from ever having ancestors. But is very easy to lose track of the inconsistency (hence the ‘paradox’). And if you are thinking about what to do, aware that you are in a one-world-history time travel scenario, and that this is your grandfather, then you will treat killing him as live practical possibility, though a dead epistemic possibility.

2.3 A Surprising Thing About Predictability-From-the-Inside

If (3) is true then either (1) or (2) must be false. Another story:

The Aspiring Prophet

An aspiring prophet has set about predicting whether an event occurs. You have access to her predictions and resolve to defy them. You (unlike Goldman) are good at defying predictions. If she predicts that the event will occur then you make sure that it does not. If she predicts that it will not occur then you make sure that it does.

What happens next? Well, the aspiring prophet fails to predict the event. Either she makes a prediction, and it is wrong, or she does not make a prediction.

More generally, if the world is counter-suggestible to an aspiring prophet's predictions then she does not accurately predict the future. Indeed, if she cannot make it the case that the world is not counter-suggestible to her predictions then she *cannot* predict the future. It doesn't matter what background epistemic resources she has. Maybe she can consult a Delphic Oracle. Maybe she can talk directly with God. No matter. If she cannot make it the case that the world is not counter-suggestible to her predictions then she cannot predict the future. Call this the *Constraint on Predictability-From-the-Inside*.

And what goes for prediction also goes for belief. If the world is counter-suggestible to the aspiring prophet's beliefs (if she believes that something will occur then it will not, if she believes that it will not then it will) then she does not have true beliefs concerning the future. And if she cannot make it the case that the world is not

counter-suggestible to her beliefs then she cannot have true beliefs about the future, no matter what epistemic resources she has.

So it was at the very least misleading for the philosopher to say that *Epistemic Determinism* follows from *Nomological Determinism*. If ‘the world’ means ‘everything’ and the candidate-knower is part of everything then this is false. The most that follows from Nomological Determinism is:

Epistemic Determinism II – A Claim About What is Knowable-From-the-Outside
Suppose the world contains a causally isolated sub-system. Suppose that you are not part of that sub-system, and you know it. It is in principle possible for you to know how that sub-system will be at any given later time, by knowing the precise way that sub-system was at any given earlier time.

To see why this is at least prima facie surprising, consider one way of fleshing out the details of our aspiring prophet’s story:

The Aspiring Prophet and the Universal Turing Machine

The prophet lives in a *computably deterministic* world. This is to say that

First, her world is governed by deterministic laws. There are no distinct, precise temporary intrinsic states of the world s_1, s_2, s_3 , number n , such that it is both consistent with the laws that the world be-in- s_1 -then-in- s_2 - n -seconds-later and consistent with the laws that the world be-in- s_1 -then-in- s_3 - n -seconds-later.

Second, all precise temporary intrinsic states of her world are finitely representable. For each precise temporary intrinsic state of the world s , there is a natural number \underline{s} that represents it.

Third, the law-governed evolution of her world can be computed. The function F such that for all states of the world s_1, s_2 , numbers n , $F(\underline{s}_1, n) = \underline{s}_2$ iff it is consistent with the laws of nature that the world be-in- s_1 -then-in- s_2 - n -seconds-later (henceforth the *evolution-function for the world*) is Turing-computable.

One fine time, t , she comes across a universal Turing machine. She is curious about the future. In an effort to find out how things will be in 10,000 seconds, she feeds the machine the program for the evolution function of her world, the number that represents the state of her world at t , and the number 10,000. But you are looking over her shoulder. For any state s , if the machine outputs \underline{s} before $t+10,000$, then you ensure that her world is not in s at $t+10,000$. If the machine outputs \underline{s} before $t+10,000$, and s is a state in which your hand is raised, then you keep your hand down at $t+10,000$. If the machine outputs \underline{s} before $t+10,000$, and s is a state in which your hand is down, then you raise your hand at $t+10,000$. The machine chugs away in just the way that it was designed to... and delivers its answer.

What happens next? Well, it is a universal Turing machine, so, given those inputs, its output must be the code for the state of the prophet's world at $t+10,000$. But it must take more than 10,000 seconds to yield that output – after all, we have said that if it yields an output in less than 10,000 seconds, then the output is not the code for the state of the prophet's world at $t+10,000$. So, on pain of inconsistency, the story must end with the universal Turing machine taking more than 10,000 seconds to perform its computation. If (1) and (3) are true then (2) is false.

Notice that, in telling this story, we said nothing directly about the workings of the machine – about how fast or slowly it operates. But it turns out that, although by embedding a universal Turing machine in a deterministic world we do not place limits on what functions it computes (any universal Turing machine can compute any Turing-computable function, given the right program), sometimes by embedding a universal Turing machine in a deterministic world we do place limits on how quickly it computes certain functions. If there are frustrators around, then universal Turing machines do not

compute evolution functions for worlds in which they are embedded faster than those worlds evolve.¹¹

2.3 A Surprising Thing About Our Ability to Predict and Control Ourselves

Michael Scriven famously argued¹² that considerations of this kind show us that human behavior is ‘essentially unpredictable’, and that the laws governing human behavior were ‘essentially unknowable’. As critics at the time¹³ were eager to point out, this was too strong. Yes, there are constraints on our predictability *from the inside* – when I have access to your predictions, and a will to defy them, then you cannot predict my behavior, no matter how clever you are. It does not follow that there are similar constraints on predictability *from the outside* – when my behavior is causally isolated

¹¹ But isn’t it possible that there be a Turing machine that performs the same computation faster? Maybe so, but that machine would not be predicting the future state of its own world. It would be predicting the future state of another world.

Here’s an analogy that at least one of the authors finds helpful: Decidably enumerate the Turing machines, then take a machine, U, such that, given input **m,n**

U gives output **0** if the m-th machine gives output **0**, given input **n**

U gives output **1** if the m-th machine gives output **>0**, given input **n**

U does not halt if the m-th machine does not halt, given input **n**

Now take a machine, NegU, such that, given input **k**

NegU gives output **1** if U gives output **0** after 1000 steps or less, given input **k,k**

NegU gives output **0** if U gives output **1** after 1000 steps or less, given input **k,k**

NegU gives output **0** if U does not halt after 1000 steps or less

Where p is NegU’s place on the list, what does U do, given input **p,p**? It gives output **0**. And it takes more than 1000 steps to complete the computation! We know this, even though we said nothing directly about how U works. A proof: Note, first, that

Correlation Given input **p,p**, U gives the output that NegU gives, given input **p**

The output must be either **1** or **0** (because NegU always gives output **1** or **0**). But it can’t be **1** (if NegU gives output **1**, given input p, then U gives output **0**, given input **p,p** – which contradicts *Correlation*). So it must be **0**. And U can’t complete the computation in 1000 steps or less (if U gives output **0** after 1000 steps or less, given input **p,p**, then Neg U gives output **1**, given input **p** – which contradicts *Correlation*). So U must complete the computation in more than 1000 steps. But isn’t it possible that there be a Turing machine, U*, that computes the same function as U, but takes fewer steps to yield output **0**, given input **p,p**? Maybe so, but then there is another machine, NegU*, such that we know that U* takes more than a 1000 steps to output a **0**, given *its* number as input.

¹² in Michael Scriven ‘An Essential Unpredictability in Human Behavior’, in Wolman and Nagel eds., *Scientific Psychology: Principles and Approaches*. Basic Books (1965).

¹³ See David Lewis, Jane Shelby Richardson, ‘Scriven on Human Unpredictability’, *Philosophical Studies*, vol. 17, no. 5 (1966), and Gil Harman, ‘Scriven on the Unknowability of Psychological Laws’, *Philosophical Studies*, vol. 18 (1967),

from your predictions then you cannot predict it, no matter how clever you are. Scriven's arguments to this conclusion were subtly fallacious.

But something interesting does follow about our predictability to ourselves.

Notice that one important class of events that may be counter-suggestible to my present belief-states is my own future behavior. Sometimes, if I believe I will do the one thing then I will do another thing, but if I believe I will do another thing then I will do the one thing.

This may be down to a positive desire to be unpredictable to myself. If I believe I will choose the chocolate ice cream then I choose the vanilla ice cream. If I believe I will choose the vanilla ice cream then I choose the chocolate ice cream.

It may be down to the ways in which my beliefs about what I will do interact with my self-conception. If I believe I will be faithful then I become complacent and unfaithful. If I believe I will be unfaithful then I become self-loathing and faithful.

It may be down to details of my behavior that are not fully within my conscious control. I am throwing darts at a board. If I believe that I will hit a double-ten, followed by a thirteen, followed by a triple-twenty, then I will not hit a double-ten, followed by a thirteen, followed by a triple-twenty. Why? Generally, when I pay too much attention to numbers then I miss them. And in this case, if I hit the double-ten, and follow it up with a thirteen, then I will be so shocked that my play is conforming to my beliefs, so unhealthily attentive to that triple-twenty, that I will miss it by a wide margin. And so it is with all fully precise beliefs about what numbers I will hit.

In any case, just so long as these aspects of my future behavior are counter-suggestible to my present beliefs then I do not have true present beliefs about my future behavior. And, just so long as I cannot render these aspects of my future behavior not-

counter-suggestible to my present beliefs then I cannot have true present beliefs about my future behavior – even if determinism is true. There is a dark pillar ahead of me, and no ingenious contraptions, no physics simulators, no oracles, will allow me to peer around it.

What goes for future-self-knowledge also goes for future-self-control. To exert present control over all aspects of my future behavior, I must know how my future behavior depends on my present behavior. I must know whether conditionals of the form:

If I do *A* now then I will do *B* later

are true. It may look as if the truth of determinism would imply that I can in principle know, for any conditional of that form, whether or not it is true. But looks are deceiving. If the truth of the conditional is counter-suggestible to my beliefs concerning the conditional, and I cannot render the truth of the conditional not-counter-suggestible to my beliefs concerning the conditional, then I cannot know whether these conditionals are true, and control my future self – even if determinism is true.

This has some bearing on the most widely discussed question concerning nomological determinism -- whether there can be free, morally responsible people in a deterministic world.

Back in the day, some philosophers¹⁴ argued that constraints on predictability-from-the-inside tell decisively in favor of compatibilism – the view that it is possible that we act freely / are morally responsible for their actions, in a nomologically deterministic world. Again, as critics at the time were eager to point out, the conclusion was too strong. No real progress can be made towards resolving the debate between

¹⁴ See Donald MacKay 'The Logical Indeterminateness of Human Choices', *The British Journal for the Philosophy of Science*, vol. 24, no. 4 (1973).

compatibilists and incompatibilists by pointing to constraints on predictability-from-the-inside . The considerations that traditionally move incompatibilists to be incompatibilists have nothing directly to do with knowledge and predictability.

Incompatibilists say that, if nomological determinism is true then

- (i) Hitler could not have acted differently.

and / or

- (ii) events outside of Hitler's control sufficiently caused his actions.

They conclude that if nomological determinism is true then Hitler was unfree and irresponsible . Maybe they are wrong, but their arguments are not in any way undermined by observing that all that follows from nomological determinism is that Hitler's actions were in principle predictable *from the outside* – by someone whose predictions were causally isolated from the actions. It may well be that his actions were only in principle predictable from the outside, and yet (i) and (ii) are true.

Nonetheless, constraints on predictability from the inside can go some way towards explaining why the truth of nomological determinism does not threaten one sort of attitude we take, an attitude that could fairly be described as 'taking ourselves to be free.' It is natural for us to divide the space of things into *things such that it is in principle possible for me to predict and control their behavior* and *things such that it is not in principle possible for me to predict and control their behavior*. Our future selves, and people whose behavior is counter-suggestible to our beliefs concerning their behavior, fall squarely in the latter category.

2.5 Back to the New Lazies

What does all this have to do with the New Lazy argument? Well, the crucial premise of that argument was that, if nomological determinism is true, then knowledge of the result of the exam is accessible to me – meaning that I now know that there’s an evidence-gathering path open to me (involving my learning more and more about past states of the world, and about the laws of nature governing the world) such that if I proceed down that path then I will be in a position to know the result of the exam. Constraints on internal predictability show us that this is not true.

I am in a position to know that if I proceed down the learning-more-and-more-about-past-states-and-laws path then I will come to know about the result of the exam, only if I know that the result of the exam is not counter-suggestible to my beliefs about the result of the exam. But I don’t know that. Indeed, I have good grounds for thinking that, if I believe that I will pass, then I will fail. So I am in position to know that if I proceed down the learning-more-and-more-about-past-states-and-laws path then I will come to know about the result of the exam, only if have grounds for thinking that if I proceed down the learning-more-and-more-about-past-states-and-laws path then I will come to know that I will fail. But I have no grounds for thinking that.

In sum: Be grateful for your counter-suggestibility. It saves you from rational sloth!