# Free Minds and Future Contingents

Roderick T. Long
Auburn University

My aim in this paper is to defend a libertarian incompatibilist account of free will against four plausible objections. First, I'll develop the objections; second, I'll offer a positive argument on behalf of libertarian incompatibilism; finally, I'll try to show how my positive account handles the objections.

## Objection 1: Free Choices Would Be Unmotivated

One celebrated objection to libertarian incompatibilism, tracing its ancestry back at least to Hume, is the claim that if actions were not necessitated by previous states of affairs they would be *random* and *uncaused*, and therefore not caused by the agent's desires, and therefore not motivated or purposeful, and therefore not actions at all – and certainly not actions over which the agent could plausibly be said to have any *control*, which is what the libertarian incompatibilist was supposed to have wanted. This objection is so familiar that I assume I need not spend more time explaining it.

## Objection 2: We Couldn't Know Whether Anyone's Choices Were Actually Free

A different (and to my mind a subtler) objection to incompatibilism about free will is that it seems to commit us to a kind of agnosticism about whether anybody actually has free will or not. If compatibilism is true, then in order to determine whether someone has acted freely, we need only apply our everyday criteria for responsibility: did the person's action express her desires? was her rationality unimpaired? was she free from deception or duress? And these are factors which we have some idea how to identify. But if incompatibilism is true, then in order to determine whether someone has acted freely, we must determine not only whether her action meets these everyday criteria, but in addition, we must determine whether her action was or was not entailed by the past history of the universe together with the laws of nature – and that is something for which we have no idea how to test.

Some incompatibilists will happily admit the truth of this point, but deny that it constitutes grounds for criticism. The incompatibility between free will and determinism, these incompatibilists will say, may be a conceptual truth, ascertainable through *a priori*

reasoning; but the question of whether free will *exists* is an empirical, scientific question – one to be settled in the neurophysiologist's laboratory rather than in the philosopher's armchair.

But surely the compatibilist objector is right to think this a very strange reply. Could we really *live* this sort of incompatibilism? Is it possible to sustain a stable attitude of agnosticism as to whether we and those around us are genuinely free agents? If incompatibilism genuinely involved this sort of deformation of our ordinary conception of ourselves and others, this would be, at the very least, a serious mark against it.

I pause to note that the second objection seems to raise a problem for the first. For the first objection implies that an action's being done freely requires its being causally *determined*, which is presumably just as hard a matter to test as its being causally *undetermined*. So *both* of these objections can't be right. Still, *either* one's being right would be enough to sink the libertarian incompatibilist's ship.

**Objection 3: Free Choices Wouldn't Fit into a Scientific Picture of the Universe**

The notion that there's something spooky or supernaturalist about free choices as the libertarian incompatibilist conceives them is a sentiment expressed frequently, but also rather vaguely. Let me try to sharpen it.

Aristotle draws a useful distinction between temporal and natural priority. Even when X and Y are simultaneous in time, X may nevertheless be prior to Y in a nontemporal sense if Y *depends* on X and not vice versa. Aristotle initially cashes out this notion of dependence in terms of asymmetric entailment: if Y entails X but not vice versa, then X is prior to Y because it can exist without Y while Y cannot exist without it. But then Aristotle adds that even in cases where X and Y entail each other, X still counts as *naturally* prior to Y if X is the reason or basis or cause of Y.[1] Whatever causes or explains something is naturally prior to what it causes or explains. What is naturally prior may be temporally prior as well, but it need not; think of Augustine's example[2] of a foot pressed into a footprint throughout all time.

---

1 *Cat.* 14 b 18-22.

2 *City of God* X. 31.

Now every physicalist accepts the core claim that all facts, including psychological facts, supervene on physical facts. But physicalism requires more than this; for mere supervenience could be accepted by dualists and idealists as well. Suppose, for example, that the physical world is constituted by the dreams of the Great Cosmic Mind. Suppose further that every alteration in the mental states of the Great Cosmic Mind automatically manifests itself in, and as, an alteration in the dream. In that case, the requirements for supervenience would be met: no mental change without a physical change! But that would hardly be a physicalist world. Physicalism requires more than a mere correlation between the mental and the physical; it requires the physical to be *in some way* the cause or reason or basis of the mental.

The asymmetry to which physicalists are committed need not involve the claim that all explanation is bottom-up or that higher-level properties are epiphenomenal. Think of my shadow, which supervenes on my shape and activity, along with how I'm spatially related to light sources and other objects. My shadow is not causally inert; I can see it and be startled by it, for example. So there can be reciprocal interaction between my shadow and its supervenience base. Nevertheless, there remains an important sense in which what my shadow does is *grounded* in my spatial relations and not vice versa. The shadow's causal powers are parasitic on the causal powers of its supervenience base. Likewise for the psychological and the physical. So let's say that X supervenes *robustly* on Y just in case X supervenes on Y, and Y is also *either identical with or naturally prior to* X.

The physicalist holds, then, that all psychological facts supervene robustly on physical facts. Does this view commit her to the more precise claim that, for any given time *t*, all psychological facts holding at t supervene robustly on physical facts holding at *t*? Strictly speaking, no, at least if *genuinely remembering at t* and *falsely remembering at t* are distinct psychological states, since these might be indistinguishable in terms of physical states holding at *t*. That, however, is true only because a psychological state holding at *t* counts as a genuine rather than a false memory only because of facts holding at times prior to *t*. Let us restrict our attention, then, to *hard* psychological facts holding at t (where a hard fact is one whose specification makes no essential reference to earlier or later times). One candidate for a hard psychological fact would be *apparent memory* – a genus of which genuine memory and false memory are species.

Is the physicalist committed to claiming that all hard psychological facts supervene robustly on temporally simultaneous hard physical facts? Perhaps she is, but I won't argue for that thesis. Instead, I shall argue for a still weaker thesis: any physicalist who rejects backward causation (which is most of them) is committed to claiming that all *indeterministic* hard psychological facts supervene robustly on temporally simultaneous hard physical facts (where a hard fact is indeterministic just in case no temporally prior hard fact is sufficient for it). Here's why. An indeterministic hard fact cannot supervene on any hard fact temporally prior to it, for if it did, then it would have a temporally prior hard fact as a sufficient condition, which is ruled out *ex hypothesi*. It can, of course, supervene on hard facts temporally posterior to it, but such supervenience cannot be robust; for if it is *logically* grounded in a future fact then it is no longer hard, and if it is *causally* grounded in a future fact then we have backwards causation.

On the physicalist view, then, any indeterministic psychological event[3] must be either a) *identical* with, or b) *naturally posterior* to, its temporally simultaneous physical supervenience base. But disjunct (a) is superfluous. Even if psychological events are identical with physical events (which a physicalist of course need not maintain), no psychological event is going to be identical with a *simple* physical event (say, the motion of a single subatomic particle); that would be enormously implausible. So a psychological event could be identical only with a *complex* physical event. And if that event is an indeterministic one, at least one of the constituent events on which it supervenes must be indeterministic as well. Now if the whole indeterministic event occurs *because* its constituents do, then the occurrence of the constituents is naturally prior to the occurrence of the whole.

This is not the only possibility, of course. In quantum physics, Bell's Nonlocality Theorem appears to show that there are synchronistically linked indeterministic events whose constituents occur because the whole does; the determination is top-down. This is a failure of robust supervenience, but does not contradict physicalism, since the non-robustly supervenient item is not itself nonphysical. Suppose that an indeterministic psychological event were identical with a Bell nonlocal quantum event. In that case, such an event would not be naturally posterior to its supervenience base. But any such neat and tidy

---

[3] I speak interchangeably of *facts* and *events* without defining either. Well, one can only bite off so much in one paper.

correspondences between psychological events and Bell nonlocal quantum events would be amazing. Barring such correspondences, the physicalist is pushed to the view that every indeterministic psychological event has naturally prior though temporally simultaneous hard sufficient conditions.

Now the libertarian believes in free choices that have no *temporally* prior hard sufficient conditions, while the physicalist believes that any such choices must have *naturally* prior but temporally simultaneous hard sufficient conditions. (Assume such choices are hard facts.) So stated, there is no conflict between the two positions. An event could quite consistently lack temporally prior hard sufficient conditions while possessing naturally prior but temporally simultaneous hard sufficient conditions. To put it another way, an event could (robustly) supervene on present physical conditions without doing so on past ones.

Nevertheless, it would be unreasonable to hold both positions. Here's why. The libertarian believes that the absence of *temporally* prior hard sufficient conditions is a precondition for a choice's being free. No one who thinks *that* could reasonably find *naturally* prior (but temporally simultaneous) hard sufficient conditions unobjectionable. After all, in both cases, one's choice is settled by factors distinct from and prior to it. If deterministic causes are a threat to freedom, robust supervenience must be so as well. For imagine saying to the libertarian: "The fact that your choices are determined by their supervenience base is no threat to freedom. They're still *your* choices; the necessitating explanation goes *through* your choice, it doesn't bypass it." If the libertarian could find such reassurances convincing in the case of robust supervenience, then she should also find them convincing in the case of causal determinism. But, *modus tollens*. What argument could make robust supervenience palatable without making causal determinism palatable as well? There is no reason to treat temporally prior sufficient conditions and naturally prior sufficient conditions differently; either both are a threat to freedom, or neither is. But the libertarian cannot say that neither is (without ceasing to be a libertarian); so she must say that both are. Determinism is just the view that the future robustly supervenes on the past; the libertarian's grounds for rejecting this will be equally good (or bad) grounds for rejecting the view that present free choices robustly supervene on *anything*.

Now those who are suspicious of narrow content may find the notion of hard psychological facts objectionable for the same reason. For example, if "meaning ain't in the head," then even an apparent memory of water will count as being that, rather than an

apparent memory of twin-water, only if it has the right causal history, and so is not a hard psychological fact after all. So why couldn't a libertarian physicalist regard psychological events, including free choices, as soft facts, and thus avoid the implication that present choices are determined by naturally prior but temporally simultaneous facts? The answer is that any view that regards present choices as soft facts will presumably have to regard such choices as consisting partly in what is happening now and partly in what has happened in the past. It would be very odd to treat present choices as consisting partly in what has not *yet* happened (except under non-essential descriptions, such as "Caesar chose to set in motion a series of events which would culminate in the destruction of the Roman republic"). But presumably if my present choice consists in past and present facts, then for the physicalist it robustly supervenes on some conjunction of past hard physical facts and present hard physical facts. If libertarians are committed (by definition) to rejecting robust supervenience on the former, and are likewise committed (by the argument I've just given) to rejecting robust supervenience on the latter, it's hard to see why robust supervenience on the conjunction of the two should fare any better.

It follows, apparently, that the libertarian incompatibilist is committed to rejecting not only determinism but also physicalism. That's not necessarily a decisive objection to the view, since physicalism remains a controversial position. But one might think it at least adds a heavy weight to the libertarian compatibilist's already considerable burden of proof.

### Objection 4: The Laws of Logic Would Fail

Our first two objections are well established in the literature; the third I haven't seen stated in precisely the form I state it, but I've suggested that it's implicit in much of the resistance to the libertarian incompatibilist position. This fourth objection, by contrast, is one that few contemporary philosophers will take seriously; but I think they should.

There is a way of thinking about possibility in ordinary language that is not captured by philosophers' talk of possible worlds. In this way of thinking, a *possible* state of affairs is one that is *reachable* from the actual world, meaning that there must be some *available route* from the actual world to this possible one. Hence while contemporary philosophers would regard a world in which England won the Battle of Hastings as a possible world, as it involves no contradiction, on the ordinary way of thinking such a scenario is only *formerly* possible, since any such possibility has since been foreclosed by what actually happened. In other words,

for contemporary philosophers a possible world need only be *internally* consistent, but on the ordinary understanding of possibility, a possible world must also be *consistent with the actual world*. This, I suggest, is the way of thinking about possibility that Aristotle, Epicurus, and the Stoics were appealing to in their treatment of future contingents. Call this conception of possibility *availability*.

Now if any available world must be consistent with the actual world, this implies that if it is a fact about the actual world that I will make a certain choice tomorrow, then any future in which I fail to make that choice is unavailable . We are then left with two options. One is to embrace the *modus ponens* and affirm that there is only one available future; this was of course the Stoic solution. The other is to embrace the *modus tollens* and insist that there is as yet no fact of the matter as to what choice I will make tomorrow, so that more than one available future is consistent with the actual world; this was of course the Aristotelean and Epicurean solution.[4]

Given this outlook, the objection to libertarianism – or indeed to any form of indeterminism – is that the indeterminist must choose the second option, and so deny truth-value to predictions of future contingent events, and so deny Bivalence. But the denial of Bivalence, so runs the objection, is absurd; therefore, indeterminism is false.

So what exactly is wrong with denying Bivalence? While many objections have been raised to the position Quine labels "Aristotle's Fantasy," I haven't seen anyone raise the criticism that the position offends against Leibniz's Law; but it seems to me the most natural objection to raise. According to Leibniz's Law, A and B are identical only if they have all the same properties. Whether or not sharing all properties is, as Leibniz also thought, sufficient for numerical identity, it is at least necessary. But according to Aristotle's Fantasy, predictions concerning future contingents are, as yet, neither true nor false. If it is not yet causally settled whether there will be a sea battle tomorrow (to use Aristotle's example), then there is as yet nothing for a prediction to correspond to or clash with – no truthmaker for either the assertion or the denial – and so any prediction's status must remain indeterminate until things turn out one way or the other.

---

[4] See Roderick T. Long, *Free Choice and Indeterminism in Aristotle and Later Antiquity*, Ph.D. dissertation, Cornell, 1992.

But these two claims seem to clash. If, as per Aristotle's Fantasy, it wasn't yet true yesterday that I would be writing about free will today, then it seems difficult to resist the inference that I have a property – writing about free will today – that my past self lacks, thus violating Leibniz's Law. While this objection to indeterminism is one that few contemporary philosophers find compelling, since I accept the availability conception of possibility it's an objection that *I*, at least, need to worry about.

Here, then, are our four objections to libertarian incompatibilism: the position is alleged to clash, first with free choices' being motivated, second with free choices' being identifiable by ordinary criteria, third with free choices' having a place in the natural order, and fourth with Leibniz's Law.

**Determinism and Deliberation**

Many arguments have been offered for libertarian incompatibilism, but I shall focus on the argument from deliberation – an argument associated most notably with Kant but arguably running back to the Greeks. According to this argument, we cannot coherently engage in deliberation unless we assume that it is *not* already settled, prior to our choice, what decision we shall make. But if we were to accept determinism, we would on the contrary have to regard this as settled. Since deliberation is an indispensable feature of our lives, we cannot accept determinism without pragmatic incoherence. Indeed, the very act of deliberating about whether to accept determinism would already commit us to rejecting it.

The obvious compatibilist response is to claim that what would render deliberation incoherent is not the assumption *that* the outcome is already settled but rather any assumption about *which way* the outcome is settled. In other words, I cannot coherently deliberate between options A and B if I believe it is already settled *that I will choose A*; but that does not show that I cannot coherently deliberate between options A and B if I believe it is already settled which one I will choose.

The example of Newcomb's Problem, however, casts doubt on this compatibilist strategy. In Newcomb's Problem, the case for choosing both boxes seems rationally *decisive*, and the case for choosing one box only also seems rationally *decisive*. Attempts to deny the decisiveness of one or the other are motivated by the assumption that they cannot *both* be decisive. It seems clear to me that they *are* both decisive, *given* the conditions laid down in

the Newcomb setup. Hence I follow George Schlesinger[5] in taking Newcomb's Problem as a pragmatic *reductio* of those conditions. What Newcomb's Problem shows is that no one can coherently take herself to be in a Newcomb's Problem situation (since if she were, she would have to take herself to have decisive reason for mutually incompatible courses of action, which is pragmatically incoherent).

The two-boxer will point out, correctly, that the opaque box has in it now, and has had in it for some time, whatever money it has in it. The money isn't hovering in a halfway state of almost-there-ness like Schrödinger's Cat; it's either there or not there, and no choice of mine can make such money magically materialise or vanish. Whether the money is there or not is already settled, and I cannot unsettle it. In other words, I must regard the presence or absence of the money as naturally prior to my choice. But on the other hand, the one-boxer will point out, correctly, that as the conditions have been set up, my choosing the opaque box only is both necessary and sufficient for the money's being in it, so if I want the money I should choose the opaque box only, like a Calvinist performing good works in order to ensure that he is *already* one of the Elect. If I can control my choice, and the money's being there or not tracks my choice, then can I not control whether the money is there? I am choosing the one-box strategy *in order to make it the case that* the money be there. In other words, I must regard the presence or absence of the money as naturally posterior to my choice. To take myself to be in a Newcomb's Problem, I must take one and the same fact to be *both* naturally prior and naturally posterior to my choice. That is why one cannot coherently take oneself to be in such a situation.

My argument assumes that when we do X for the sake of Y, we must regard X as a cause of Y and so as naturally prior to Y. Richard Taylor denies this:

> [T]here are certain purposeful actions, the ends or goals of which … *precede* those actions. For example, if it is one's purpose to flex a certain arm muscle, the only way he can accomplish this, normally, is by moving his arm. The arm is *caused* to move by the motion of the muscle, but the arm is moved *in order* to move the muscle, not vice versa. The motion of the muscle does not follow upon the motion of the arm, however … [but] precedes it slightly. Or consider nerve impulses. Part of the cause of the motion of a man's arm, when he moves it, is certainly a certain nerve impulse from his brain, but it is

[5] George Schlesinger, "The Unpredictability of Free Choices," *British Journal for the Philosophy of Science* 25 (1974): 209-21; cf. Schlesinger, *Aspects of Time* (Indianapolis: Hackett, 1980), pp. 79, 144.

> false that he moves his arm *by means* of the nerve impulse. It is the other way around. Thus, if a man learns of these nerve impulses – and many go to their graves without ever suspecting they exist – and then has some occasion to produce one of them – perhaps for the purposes of some experiment in physiological psychology – he can do so only by moving his arm. The purposeful action in such a case – namely, the man's moving his arm – is the means to a certain end – the production of a certain nerve impulse – which actually precedes that purposeful action in time.[6]

Taylor finds this result "frightfully puzzling," and well he should. His example strikes me as simply a *reductio* of his view that basic actions are bits of overt behaviour, rather than volitions. Without that assumption, the paradox dissolves.

But what is it about the conditions of the Newcomb setup that renders pragmatically incoherent the assumption that we are in one? If it is only knowing *which* option is settled that renders deliberation incoherent, then the incoherence of Newcomb's Problem is unexplained, since the agent does not know which option is settled. What explains the incoherence, I suggest, is the mere assumption that some option is settled. Whether the money is in the opaque box is already settled, since the predictor has long since decided whether to put in the money or not; but the presence of the money is sufficient (not *causally* sufficient, but sufficient nonetheless, as the Newcomb story goes) for my choosing the one-box option, and its absence is sufficient for my choosing the two-box option, so it *is* already settled what option I will choose; and *this* is the crucial feature that leads to the paradox.

The Newcomb setup as it stands includes *three* odd features: a) the fact that my choice is already settled, b) the fact that the predictor knows which way it's settled, and c) the fact that I know that (though not what) the predictor knows. It's easy to be distracted by (b) and (c), and to take them, rather than (a), as generating the paradox. But (b) and (c) are accidental and eliminable features of the setup; for we can construct a version of Newcomb's Problem without any predictor at all.

Let us define C as the disjunction of those conditions that might possibly hold at $t_1$ the holding of any of which would be causally sufficient for my $\phi$-ing at $t_2$. (We can think of choosing just the opaque box rather than both as a possible instance of $\phi$-ing, and we can think of whatever leads the predictor to predict my one-box strategy and so to place the money in the box as a possible instance of C. But we needn't; $\phi$-ing and C can be anything,

---

6 *Action and Purpose* (Englewood Cliffs: Prentice-Hall, 1966), pp. 194-195.

so long as they stand in the causal relation described.) If causal determinism is true, then the occurrence of C at $t_1$ is causally necessary for my ϕ-ing at $t_2$; if not, not.

Now there seem to be very few limits to what human beings might conceivably have preferences about; so I am surely calling for a harmlessly modest assumption if I ask us to assume that I might have preferences rank-ordered as follows.

1. C holds at $t_1$ and I do not ϕ at $t_2$
2. C holds at $t_1$ and I ϕ at $t_2$
3. C does not hold at $t_1$ and I do not ϕ at $t_2$
4. C does not hold at $t_1$ and I ϕ at $t_2$

Now if, at $t_2$, it is outside my control whether C holds at $t_1$ – which it must be, since it is settled and past – then the rational thing for me to do, clearly, is *not* to ϕ at $t_2$. For if C does hold at $t_1$, then not ϕ-ing at $t_2$ will get me option 1 over option 2; and if C does not hold at $t_1$, then not ϕ-ing at $t_2$ will get me option 3 over option 4. However, if I accept determinism as true, then the opposite reasoning also holds; for in that case the only possible options are 2 and 3. If so, then C's occurrence at $t_1$ is necessary for my ϕ-ing at $t_2$, and so my ϕ-ing at $t_2$ is sufficient for C's occurrence at $t_1$; and since I cannot help regarding the choice between ϕ-ing and not ϕ-ing as up to me, I must regard it as in my power to guarantee, by my present ϕ-ing, the holding of C in the past, and so it is rational for me to ϕ. (That, incidentally, is why David Lewis is wrong to assimilate Newcomb's Problem to the Prisoner's Dilemma.)[7]

We may phrase the same argument slightly differently: Let D1 be the disjunction of all those possible pasts that, when conjoined with the actual laws of nature, entail my ϕ-ing at *t*; and let D2 be the disjunction of all those possible pasts that, when conjoined with the actual laws of nature, entail my not ϕ-ing at *t*.

Now suppose I believe it is causally determined whether or not I shall ϕ at *t*. In that case, I am committed to assuming that the actual past belongs either to D1 or to D2.

Suppose further that I have, as I might, the following preference ordering:

A. The actual past belongs to D1, and I do not ϕ at *t*

---

7 David Lewis, "Prisoners' Dilemma Is a Newcomb Problem," *Philosophy & Public Affairs* 8 (1979), pp. 235-240.

B. The actual past belongs to D1, and I ϕ at *t*
C. The actual past belongs to D2, and I do not ϕ at *t*
D. The actual past belongs to D2, and I ϕ at *t*

*Ex hypothesi* I can deliberate about whether or not to ϕ at *t*. But since I cannot change the past, it makes no sense for me to deliberate about what the past is to be; I must take it as already settled whether the actual past belongs to D1 or D2. If the actual past belongs to D1, the only options are A and B, so I ought to refrain from ϕ-ing at *t*; if the actual past belongs to D2, the only options are C and D, so once again I ought to refrain from ϕ-ing at *t*. In either case, then, it is rational for me to refrain from ϕ-ing at *t* whatever the past is. But if I also take it to be causally determined whether or not I shall ϕ at *t*, then the only possible options are B and C. Hence the rational thing for me to do is to ϕ at *t*.

The assumption that it is causally determined whether or not I shall ϕ at *t*, conjoined with the assumption that I have a certain preference ordering which I certainly might have, entails both that the rational thing for me to do is to ϕ at *t* and that the rational thing for me to do is to refrain from ϕ-ing at *t*. As a deliberative agent I cannot accept that there is nothing that I rationally ought to do. If the conjunction of two assumptions commits me to a contradiction, I must reject one of the assumptions. Since the relevant preference ordering is one that I can freely adopt at any time – for the purpose of refuting determinism, for example – I cannot coherently reject the assumption that I might have that preference ordering. Instead, then, I must reject the assumption that it is causally determined whether or not I shall ϕ at *t*. In short, I must affirm my own libertarian freedom *a priori*.

To put the point slightly differently: compatibilists and incompatibilists often disagree as to whether coherent deliberation requires the assumption that alternatives are *causally* open, or only the assumption that they are *epistemically* open. But if the epistemic assumption were all that's required, then we should be able to deliberate coherently while taking ourselves to be in a Newcomb situation. Ergo, *modus tollens*.

Objection: The fact that we cannot coherently regard our choices as determined does not entail that they are not so; there is no incompatibility between a claim's being *true* and its endorsement being *pragmatically incoherent*. Reply: That is quite true; although determinism is pragmatically incoherent, its truth remains logically possible. But that fact cannot count as a

reason to take determinism seriously after all; for to say that the logical possibility of determinism's truth gives us reason to regard it as a live theoretical option is to say that the logical possibility of determinism's truth is enough to render the endorsement of determinism coherent. (If a claim cannot be coherently endorsed, then it is not a live theoretical option.) And if the logical possibility of determinism's truth *does* render the endorsement of determinism coherent, then the possibility claim and the incoherence claim are not compatible after all; the conclusion of the argument contradicts the premise it starts from. The determinist cannot have it both ways.

That's my brief for libertarian free will. Now let's return to the objections.

**Reply to Objection 1**

The first objection was that an action's being motivated consists in its being causally determined by the agent's motives – as though the motivation of an action is a matter of what happens *before* the action.[8] But this would be a distortion of our understanding of motives. Suppose I'm crawling in the desert, dying of thirst, and suddenly a friendly sheikh pops up over the next dune and offers me a canteen of water, saying "I'll give you a million dollars if you drink this." I do indeed eagerly accept the water – but not because of the money, which at that moment I am too thirst-crazed to care about. Yet what makes my choice to drink the water an act motivated by *thirst* rather than by *avarice*? We could talk about what antecedent mental state impels my choice, but it seems to me that what's more important is something internal to the choice's structure. When I choose the water, I choose it *as* a satisfier of my thirst, rather than as a satisfier of my avarice. (In Kantian terms, reference to thirst is part of the maxim of my action, as it were, while reference to avarice is not.) What makes a choice count as motivated by one motive rather than another has less to do with the motive's antecedent role in triggering the choice than with its internal role in constituting and specifying that choice.

The moral is that a choice need not be antecedently necessitated by a pre-existing motive X in order to count as motivated by X. A choice, however caused, counts as motivated by motive X so long as a reference to motive X is built into the *internal* structure of that choice,

[8] Cf. Wittgenstein: "The causal connection between speech and action is an external relation, whereas we need an internal one." (*PR* VI. 64.)

as a *constituent* of it, whether that choice is causally necessitated or not. Thus reason and desire are to be regarded as different *aspects* of the soul, rather than as separate homunculi within it. This is presumably why Aquinas treats reasons as formal rather than efficient causes of volition.[9]

But the determinist can fairly object that motives must play a causal role as well if motivated actions are to be intelligible. If a free choice, with a built-in motive M, simply occurs at *t* without the agent having had any inclination toward M prior to *t*, the choice does seem unintelligible. It seems unintelligible for me to murder Eric at *t*, out of hatred for him, if my hatred for Eric did not pre-exist my choice to murder him. As Wittgenstein points out, there are some properties that nothing could *count* as having except in virtue of a wider temporal context than the immediate moment.[10] In the same vein, arguably nothing could count as an act done out of hatred unless the hatred pre-existed the act. Hence intelligible behavior must be at least to some extent predictable on the basis of the agent's prior motives.

---

[9] "Is choice an act of the will, or of reason? … Choice is neither appetite by itself nor deliberation alone, but something composed of these – for just as we say that a living thing is composed of soul and body, yet is neither body by itself nor soul alone, but is both, so it is with choice. … But whenever two things come together to constitute some one thing, one of them is formal with respect to the other. … It is evident that reason precedes the will in some way, and gives order to its act – insofar, i.e., as the will tends to its object in accordance with the ordering of reason, inasmuch as the cognitive power presents to the appetitive its object. Therefore, that act whereby the will tends toward something that is put forward as good, from the fact that it is ordered to the end by reason, belongs materially to the will but formally to reason." (Thomas Aquinas, *Summa Theologiae* II. 1. 13. 1.)

Philippa Foot, too, sees that doing something for a motive is more a matter of *how* one does it than of what triggered the action: "[Some think] that when a man does something meaning to do it, he does what he wants to do, and so his action is determined by his desire. But to do something meaning to do it is to do it in a certain way, not to do it as the result of the operation of a causal law." ("Free Will as Involving Determinism," pp. 64-65; in Foot, *Virtues and Vices and Other Essays in Moral Philosophy* (Berkeley: University of California Press, 1981), pp. 62-73.)

[10] "Could someone have a feeling of ardent love or hope for the space of one second – *no matter what* preceded or followed this second? What is happening now has significance – in these surroundings." (*PI* I. 583.) "Why does it sound queer to say: 'For a second he felt deep grief'? Only because it so seldom happens?" (*PI* II. 1.) "The application of the concept 'following a rule' presupposes a custom. Hence it would be nonsense to say: just once in the history of the world someone followed a rule (or a signpost; played a game, uttered a sentence, or understood one; and so on)." (*RFM* VI. 21.)

This is in effect Wittgenstein's development of the Aristotelean idea that no condition lasting only for a moment could count as happiness, since "one swallow does not make a spring"; think also of conditions like *health*, *peace*, and *commitment*. (One can see this as the flipside of the Kantian idea that lying depends for its intelligibility on the presupposition of a general practice of truth-telling, so that universal lying is impossible. Some things by their nature can't be exceptional or momentary; other things by their nature can't *but* be exceptional or momentary.)

But acknowledging this need not imply any concession to determinism, for motives can play antecedent causal roles without being sufficient conditions. They can, for example, be both *necessary* conditions and *probabilifying* ones. Choices are something we *do* with the motives we already have. And if the determinist objects that we don't really count as being in control of our actions if our motives are only contingently related to the choices they motivate, we can appeal once again to motives as constituents of choices. As constituents, motives necessitate choices but do not precede them; as causes, motives precede choices but do not necessitate them.

May the determinist retreat to the position that what's settled is not any particular actions, but only a general tendency? But such a reply would be ambiguous between two possibilities:

Possibility One: Suppose I have an innate tendency to choose X over Y. Now that might mean that although I won't always choose X over Y, I will do so in, say, 75% of future instances. But that too seems incompatible with the deliberation argument; if it's already settled that I will choose X over Y 75% of the time, then I'm not really free to plan my future conduct. One can't regard one's future actions as the outcome of a dice throw one sits back and awaits.

Possibility Two: Now as the 16th-century Aristotelean Pietro Pomponazzi pointed out,[11] there is certainly a *sense* of "tendency" that is perfectly compatible with the deliberation argument: one might say "your present pattern of activity is such that, if it continues, you will behave in such-and-such a way."[12] Tendencies so understood pose no obstacle to free choice because it's up to the person whether a given pattern of activity in fact continues. And such a tendency might very well be innate in the sense that it is the tendency with which a person starts out. But it can't be innate in the stronger sense of being a tendency that person carries with her throughout life; for once the person changes her pattern of activity, then it is no longer true of her that "her present pattern of activity is such that ...." – and so she no longer has the tendency.[13]

---

11 *De fato, libero arbitrio, et prædestinatione* (1520).

12 cf. Wittgenstein, *RPP* I. 61, I. 219.

13 Does this mean that free will is incompatible with the existence of psychological dispositions? By no means. There are dispositions for us to have certain desires, and for acts to have certain psychic benefits and psychic costs. But those are *thymological* dispositions. What free will rules out are *praxeological* dispositions. The

What the Wittgensteinian considerations show is that the connection between having antecedent desires and acting on them, while less direct and necessary than in the case of constitutive motives, is still partly conceptual rather than purely causal. In short, the possession of a given desire is not logically compatible with *any and all* patterns of action, but only with some. Hence *we can influence what desires we count as having, to the extent that we freely determine which patterns of action we instantiate.* It follows that habituation – the fact that "use almost can change the stamp of nature" by making actions easier through repetition – is no mere empirical datum but a logically necessary feature of all free agency. It also follows that the extent to which genetic endowments constrain people's destinies is severely limited.[14]

**Reply to Objection 2**

The second objection was that if determinism is made a necessary condition of free choices, our ordinary methods of determining whether a choice was free will be undermined. But here the libertarian's apparent predicament arises from the conjunction of two theses: that the incompatibility of an action's being done freely with its being causally determined is a conceptual truth verified by philosophical means, and that the question of whether causally undetermined actions actually occur is an empirical matter best left to scientists. My version of libertarianism, however, rejects the second thesis; I've argued for the causally undetermined nature of our choices as a necessary *a priori* presupposition, not as an empirical claim. So isn't my version of libertarianism immune to the second objection?

Well, suppose my argument from deliberation is successful. Even so, should it satisfy the compatibilist objector? I think not. For all this argument shows is that I cannot

---

most the sociobiologist can claim, then, is that some thymological dispositions are genetically determined; and even here she must tread carefully. For desires, as we've seen, must have *some* fairly regular expression in action in order to count as desires at all; and to the extent that such expression is subject to volitional control, even our thymological dispositions fall to some degree under the will's sway. So it is that "use almost can change the stamp of nature"; the power of habituation, far from being inconsistent with free will, is in fact entailed by it.

[14] Now that the distinction between antecedent and constitutive motives is in place, we can also notice that it is not exhaustive. Suppose that as I type these words I'm feeling a slight itch, which of course is a motive for scratching. But the itch is *quite* mild, and I'm absorbed in what I'm typing and don't want to be distracted, so I just keep on typing rather than scratch. Now my itch isn't *antecedent* to my action of typing; it's temporally concurrent with it. But it's not a *constitutive* motive of my typing, or indeed of any act that I'm performing right now; it's just there, not yet bothersome enough to provoke me to action. And perhaps it will end up going away before I ever get around to doing anything about it. This itch is a motive that never makes it past the threshold of action (though if my itches *never* provoked me to actual scratching, their status as genuine itches would become doubtful).

coherently take *myself* to be causally determined. But it does not show that I cannot regard *you* as causally determined. All it shows is that it would be irrational for *you* to believe that you are causally determined. The statement "I cannot coherently take my actions to be causally determined, but they are causally determined anyway" is Moore-Paradoxical, or something like Moore-Paradoxical; but it loses its Moore-Paradoxicality when transposed into the second or third person. Thus the premise "S must on pain of irrationality reject the assumption that S's actions are causally determined" can license the conclusion "S's actions are not causally determined" only *for S*, not for anybody else. We seem to be left with a kind of libertarian solipsism; we can ascertain our own incompatibilist freedom *a priori*, but this *a priori* method cannot be extended to other people, and the methods that *are* available to us – our familiar everyday criteria for responsibility – seem to allow us to identify in others only the compatibilist variety of freedom. Incompatibilist agnosticism has been defeated in the first-person case, *and there only*. (Actually, the problem is worse yet: the *a priori* argument licenses confidence only in the freedom of my prospective choices, not of my past ones; all that I have been and done slides over the horizon of libertarian solipsism into the inscrutable realm of other people's agency.)

How, then, can the epistemic gap from first-person freedom to third-person freedom be bridged? (As we'll see, this may not be quite the right way to state the question, but it will do for now.) Let's start by eliminating a couple of possible answers that I think will not give us what we are looking for.

First, the libertarian might say: "Look, we've just established that I can know, on *a priori* grounds, that *I* have libertarian free will. Now what are the odds that I'm the only human being who has it? Although I can't prove that others are free agents the way I can prove that I am one, once I'm allowed to take my own freedom as given, isn't it more likely that libertarian free will is a common human trait?"

I'm not certain that this is a bad argument, but I suspect it may be. As an *inductive* argument it would seem to be inadequate, generalising as it does from a single case. It is more plausibly read as an *abductive* argument, reasoning that the probability of my having a given quality is higher if the quality is a common human trait than if it is not. Now this would work well enough, I think, in the case of empirically ascertained qualities. But if my having libertarian free will is *a priori* (for me), then its contrary is inconceivable (for me); and so the conditional probability of my having libertarian free will remains exactly the same for

me, namely 100%, whether or not anyone else has free will. If this is correct, then I cannot reason abductively from my own libertarian freedom to that of others.

A different incompatibilist response might go as follows: "This libertarian solipsism you threaten me with is an idle bogey. Perhaps on my view it's logically possible that everybody else should lack free will; but that's not a hypothesis I can seriously entertain. In living as a human being among other human beings, I just *do* regard them as free agents in the same way that I am, and that's the end of it."

Now in a sense I think that this sort of response is quite legitimate. But the compatibilist objector can protest, with some show of justice, that the incompatibilist has no right to offer it. For the response takes our everyday criteria for free will as *good enough*; but how *could* they be good enough, if free will involves not only satisfying those everyday criteria but also not being causally necessitated by the past, a separate criterion that seems to have nothing to do with the everyday ones? Finding that a person's action expressed her desires, that her rationality was unimpaired, that she was not subject to deception or duress, and so on (pick your own favourite everyday criteria), seems utterly *irrelevant* as evidence on the question of whether her action was causally determined – by contrast with the way that, for example, our everyday criteria for water, while not identical with the chemical criteria for $H_2O$, *are* relevant as evidence for the presence of $H_2O$.

In the case of free will, the incompatibilist criterion simply seems epistemically disconnected from the everyday criteria; to treat the everyday criteria as good enough is to treat the incompatibilist criterion as irrelevant *in practice*. But if it's irrelevant in practice, why insist on it in theory? Or, contrariwise, if the incompatibilist does continue to insist on it in theory, hadn't she better bring her practice in line, in which case her blithe confidence in the everyday criteria will then seem unwarranted?

Suppose I defined an *invisicorn* as a horselike creature with an invisible horn. And then suppose that every time I saw a horselike creature I called it an invisicorn, since it satisfied all the criteria for being an invisicorn except the part about the horn. An objector might say, "Look, if by 'invisicorn' you just mean 'horse', which is how in practice you seem to be applying the term, then why not stop the pretense about these things having invisible horns? Or if instead you're serious about the difference between horses and invisicorns, then what entitles you to call these creatures invisicorns when the only criteria you've actually tested for don't distinguish between invisicorns and horses?" I might reply "Peddle your skepticism

elsewhere, sophist! The criteria I'm using are good enough for practical purposes." But the rejoinder would hardly be a convincing one; and our incompatibilist seems to be saddled with a similar dilemma.

What would count as a solution to the incompatibilist's problem? Well, suppose it were part of the *concept* of free will not only that determinism is incompatible with it but *also* that our everyday criteria are evidence for it. Not *decisive* evidence, of course – if our everyday criteria were decisive evidence for free will then compatibilism would be true – but *prima facie* evidence. In that case, by accepting the concept of free will we would be committing ourselves *both* to acknowledging its indeterministic character *and* to taking other people to possess it – so long as those other people satisfy the everyday criteria and no countervailing evidence arises. Wouldn't this solve the incompatibilist's problem?

It might seem that it wouldn't; for the analogous solution in the case of the invisicorn is pretty clearly unacceptable. Suppose I said, "Well, it's part of the concept of an invisicorn not only that it has a horn but also that being horselike is *prima facie* evidence for something's being an invisicorn. So by accepting the concept of *invisicorn* you're committing yourself both to acknowledging that invisicorns have invisible horns and to taking any horselike creature to be an invisicorn until proven otherwise." In that case the obvious response would be, "If that's what you mean by 'invisicorn' then I *don't* accept the concept; it's an illegitimate concept."

There's an important asymmetry between libertarian free will and the invisicorn, however. We are free to reject the concept of invisicorns. But we are not likewise free to reject the concept of libertarian free will, since we necessarily apply it to ourselves when we deliberate.

How might it be shown, though, that the *prima facie* reliability of our everyday criteria for free agency is part of the concept, so that in attributing it to ourselves on the basis of an *a priori* argument, we are *eo ipso* committed to attributing it to others, at least presumptively, on the basis of *a posteriori* observable behaviour?

In this connection I would like to appeal to two principles. The first is that *the ability to apply a concept is part of having the concept.* If you can't apply a concept – that is, if you lack the

ability to recognise and identify instances of the concept – then you don't even possess the concept. We may call this the Wittgenstein-Rand Principle.[15]

Why should we accept the Wittgenstein-Rand Principle? Well, suppose it were false; suppose, that is, that I could possess a concept without having the ability to apply it. Let's say I have the concept *tiger*, but in practice I can't tell a tiger from a tidal pool. Yet, allegedly, this inert psychical lump lodged in my head *is* about tigers and not about tidal pools. But what is it about whatever I've got in my head that *makes* it the concept *tiger* and not the concept *tidal pool*? Perhaps, as many of our philosophical forebears were tempted to think, it *resembles* a tiger. Well, suppose it does. How does this help? A little orange-and-black image in my head might refer to all tigers, or just the orange-and-black ones, or perhaps some one individual tiger. Nothing in the image itself settles the matter; it all depends how I'm disposed to *apply* the image.[16] Unless I can reliably (not infallibly, but reliably) identify tigers when I meet them, I don't count as having the concept *tiger*. Hence criteria for applying a concept are built into the concept itself. (This is the grain of truth in verificationism – only a grain, because nothing in this argument requires the criteria to be decisive rather than presumptive.)

How is all this relevant to the problem of libertarian solipsism? Well, if in deliberation we necessarily attribute free agency to ourselves, then we must take ourselves to possess the

---

15 Wittgenstein writes: "What use of a word characterizes that word as being a negation? … It is not a question of our first *having* negation, and then asking what logical laws must hold of it in order for us to be able to use it in a certain way. The point is that using it in a certain way is what we mean by negating with it." (*Lectures on the Foundations of Mathematics* XX, p. 191) "We say: if a child has mastered language – and hence its application – it must know the meaning of words. It must, for example, be able to attach the name of its colour to a white, black, red or blue object without the occurrence of any doubt." (*On Certainty* § 522)

Analogously, Rand writes: "In order to think at all, man must be able to perform this cycle: he must know how to see an abstraction in the concrete and the concrete in an abstraction, and always relate one to the other. He must be able to derive an abstraction from the concrete … then be able to apply the abstraction …. Example: a man who has understood and accepted the abstract principle of unalienable individual rights cannot then go about advocating compulsory labor conscription …. Those who do have not performed either part of the cycle: neither the abstraction nor the translating of the abstraction into the concrete. The cycle *is unbreakable*; no part of it can be of any use, until and unless the cycle is completed …. A broken electric circuit does not function in the separate parts; it must be unbroken or there is no current …." (*Journals*, p. 481)

16 As Wittgenstein writes: "The sentence 'I imagine so-and-so' is not a *description* of a picture before my mind's eye. Ask yourself: do you recognise him from the picture before your mind's eye? Would you say: 'I see a man with white hair, etc., I suppose I'm imagining N but perhaps it's only someone who looks very much like him'?" (*Philosophical Occasions* XIV, p. 455) "I cannot know what he's planning in his heart. But suppose he always wrote out his plans; of what importance would they be? If, for example, he never acted on them. … Perhaps someone will say: Well, then they really aren't plans. But then neither would they be plans if they were *inside* him, and looking into him would do us no good." (*Last Writings on Philosophy of Psychology* I. 234-235)

concept of free agency. But we cannot possess the concept of free agency unless we possess criteria for recognising actual instances of free agency; therefore we do possess such criteria.

But does this dissolve the solipsistic predicament? Or can it be objected that our ability to apply the concept *in our own case* is enough to give us title to the concept? Might the *a priori* argument by which we established our own free agency constitute the concept's criteria of application, leaving us just as much in the dark as ever when it comes to the free agency of others?

I mentioned that I wished to invoke two principles. The first was that the ability to apply a concept is part of having the concept. Now it is time for the second: that *in the case of a concept susceptible of general instantiation, the ability to apply the concept across a suitable range of cases is part of having the concept*. We might call this the Socratic Principle.[17]

Why accept the Socratic Principle? I think it is an inescapable corollary of the Wittgenstein-Rand Principle. When I attribute free agency to myself, "free agency" is not supposed to be a *proper name* for some essentially private possession of my own. What I attribute to myself is something that others could in principle possess; *free agency* is a concept susceptible of general instantiation. But what makes it so? What is it about my concept of free agency that determines whether it is a general concept or a proper name? Just as in the earlier case of the tiger-image, it must be the manner in which I am disposed to apply it. It follows that if I am unable to apply a concept except in my own case, then what I've got is *not a general concept*. Part of possessing the concept of free agency is having the ability to recognise instances of free agency not only in one's own actions but also in the actions of others. But I cannot use the *a priori* argument from deliberation to identify free agency in others, so the concept's built-in criteria of application must include other markers as well. In attributing libertarian free agency to myself on *a priori* grounds, I have thereby committed myself to accepting public, *a posteriori* criteria for attributing it to others.[18]

---

[17] In Plato's *Laches*, Socrates argues that no one counts as possessing knowledge of good and evil if he can identify only future goods and evils, not present or past ones; and in Plato's *Ion*, Socrates argues that no one counts as possessing knowledge of poetic craft if he can identify effective poetic devices only when they occur in the works of Homer and not in those of other poets.

[18] Notice that we have arrived at the *conclusion* of Wittgenstein's private-language argument without having to make use of the controversial premise that whatever measures up to a standard must be capable of failing to measure up to it.

The incompatibilist, then, is not only *entitled* but actually *required* to accept the satisfaction of certain public criteria as *prima facie* evidence for free agency. But why, it may be asked, should we take these criteria to be identical with our everyday criteria? The answer is that the meaning of *free agency* depends on how we are *actually* disposed to apply the concept. But the everyday criteria just are the criteria we are disposed to use. (Could we have different criteria? Not for *this* concept, because this concept is individuated, in part, by our *actual* practice of applying it.) We necessarily attribute libertarian free agency to ourselves, on *a priori* grounds; but we could not do so unless we also possessed criteria for attributing it to others, since without such criteria we could not *possess* the concept we employ in the self-attribution. Thus we never could have been in the predicament of libertarian solipsism in the first place. *Free your mind, and the rest will follow.*

**Reply to Objection 3**

The third objection was that if libertarians reject the determination of free choices *by the past*, they should on the same grounds reject the determination of free choices *by those choices' supervenience base*, and that this would implausibly render free will incompatible not only with determinism but also with physicalism.

To begin with, it's true that the libertarian should find robust supervenience problematic for the same reason she finds determinism problematic. Recall the revised Newcomb's Problem argument that shows we cannot coherently regard our free choices as covarying with prior events or facts already settled. The same paradox results if we regard the forbidden covarying fact, not as a fact about the *past*, but as a fact about the *present* supervenience base of my free actions. Let P be the disjunction of all those physical facts that might possibly hold now, upon the holding of any of which my $\phi$-ing now would robustly supervene. Since I can certainly have preferences concerning whether P holds, precisely the same reasoning applies as before. Assume the following rank-ordering of preferences:

1. P holds at *t* and I do not $\phi$ at *t*
2. P holds at *t* and I $\phi$ at *t*
3. P does not hold at *t* and I do not $\phi$ at *t*
4. P does not hold at *t* and I $\phi$ at *t*

Now if, at *t*, it is outside my control whether P holds at *t* – which it must be, if my freely ϕ-ing supervenes *robustly* on its microphysical correlates – then the rational thing for me to do, clearly, is *not* to ϕ at *t*. For if P does hold at *t*, then not ϕ-ing at *t* will get me option 1 over option 2; and if P does not hold at *t*, then not ϕ-ing at *t* will get me option 3 over option 4. However, if I accept robust supervenience as true, then the opposite reasoning also holds; for in that case the only possible options are 2 and 3. If so, then P's occurrence at *t* is necessary for my ϕ-ing at *t*, and so my ϕ-ing at *t* is sufficient for C's occurrence at *t*; and since I cannot help regarding the choice between ϕ-ing and not ϕ-ing as up to me, I must regard it as in my power to guarantee, by my ϕ-ing, the holding of my ϕ-ing;s supervenience base P, and so it is rational for me to ϕ.

Let us generalize: If I make a certain choice *in order* to make it the case that some fact hold, then I am necessarily taking my choice to cause or bring about that that fact holds, and so must regard that fact as naturally posterior to my choice. But for *any* fact the settling of which settles whether I make a certain choice, I could conceivably have a preference concerning whether that fact holds, and so could make a choice in order that the fact hold. Hence I cannot coherently regard any such fact as naturally prior to my choice. But if either determinism or robust supervenience is true, then there are facts that are both a) naturally prior to my choice, and b) such that settling those facts settles whether I make that choice. Hence determinism and robust supervenience are both pragmatically incoherent. For any factor that co-varies with my free choices, I must be able to determine whether that factor holds or not – and that includes those microphysical events on which my choices supervene.

Does this mean that libertarian free will is incompatible with physicalism? I won't attempt to answer that precise question, since there is little consensus on what exactly physicalism is. Let's consider a better question: does the denial of robust supervenience, and in particular the claim that our free choices explain their supervenience bases rather than vice versa, entail some spooky or supernaturalist thesis about the ability of free agents to act directly on their microphysical constituents? I claim that the answer to this question is *no*. The way that a free agent brings about the microphysical correlates of her actions is simply by performing the actions that supervene on those correlates.

The libertarian may appear to face a dilemma, however. In performing a free action, does the agent cause that action's microphysical supervenience base to be otherwise than it

would have been according to the laws of nature? If the libertarian says yes, then she seems to be landed with something objectionably spooky. If the libertarian says no, then it's hard to see how robust supervenience has been avoided.

My reply is that the question makes sense only in a determinist framework. If we assume an indeterministic supervenience base – as we must, since indeterministic events can't supervene on deterministic ones – then it makes no sense to ask whether the agent causes that base to be otherwise than it would have been according to the laws of nature, for those laws will then be probabilistic and consistent with more than one outcome, so that there's no such thing as *the* way the base "would have been." And after all, all that the physical laws determine is the *probabilities* of certain events, and the acts of will *don't alter those probabilities at all*; the act of will doesn't make the microphysical event *more probable*, it just makes the event *happen*.

But how so? Why say that the free choice explains its supervenience base and not vice versa? Well, the supervenience base is an indeterministic microphysical event that has no explanation. (That is, it has no *specific* explanation, *i.e.*, no explanation of why it *rather than one of the other possibilities* happened. It of course has a *generic* explanation, in that it is one of the many realisations possible under the circumstances to the entities involved.) But while the free action is also an indeterministic event, it *does* have an explanation (and a specific one, not just a generic one), namely the action's (constitutive) *motive*. Hence if we consider the entire complex of the free action together with its microphysical supervenience base, this complex *has an explanation* in respect of its free-action aspect considered in its own right, but *has no explanation* in respect of its microphysical aspect considered in its own right. Hence the free-action aspect, not the microphysical aspect, is explanatorily dominant. I take this state of affairs to be inconsistent with robust supervenience. Whether this state of affairs is consistent or inconsistent with physicalism may be a terminological matter; but in any case, nothing spooky or supernatural appears to be involved.

**Reply to Objection 4**

The fourth objection was that Aristotle's Fantasy, to which I've argued the libertarian is committed, is inconsistent with Leibniz's Law, since if some prediction about me at $t_2$ lacks truth-value at $t_1$, then I will have a time-indexed property at $t_2$ that I lack at $t_1$, in defiance of the indiscernibility of identicals.

The problem of reconciling Leibniz's Law with changing truth-values is, I shall suggest, simply a trickier version of the problem of reconciling Leibniz's Law with change of any sort; so let's consider that problem first. The claim that numerical identicals must share all their properties conflicts, at first sight, with the reality of change. Are we not constantly losing old properties and gaining new ones? Do we not survive such changes with our numerical identity intact? How then can it be said that I share all my properties with my past self? This is the problem that was worrying Heracleitus when he asked whether we can step into the same river twice (a question he famously answered no, and less famously answered yes).

Of course such problems are ordinarily solved by interpreting Leibniz's Law in such a way that it only applies to *time-indexed* properties. So although my 2006 self might seem to differ from my 1996 self in being an Alabama resident, thereby imperiling Leibniz's Law, my 1996 self and my 2006 self – so the argument goes – do not differ with respect to being-an-Alabama-resident-in-2006 (they both have that property), nor do they differ with respect to being-an-Alabama-resident-in-1996 (they both lack that property). Thus it is customary to reconcile Leibniz's Law with survival through change by reinterpreting Leibniz's Law so as to restrict its application to time-indexed properties like being-an-Alabama-resident-in-2006, rather than allowing it to be applied to apparently non-time-indexed properties like being an Alabama resident *tout court*. (Another way to put it, perhaps, is that Leibniz's Law is rescued by denying that there *is* any such property as being an Alabama resident *tout court*; talk of being an Alabama resident must always be interpreted as talk of being an Alabama resident *at* some time or other.)

While this solution solves the problem of Heracleitus' river, it doesn't initially seem to help us with Aristotle's sea battle. For if in 1996 it wasn't yet settled that I would be an Alabama resident in 2006, then it seems I *didn't* have *even the time-indexed property* back then; not only did I not have, in 1996, the property of being an Alabama resident *tout court* (if there is such a property), but I also didn't have, in 1996, the property of being-an-Alabama-resident-in-2006. But I do have that property now. Hence if Aristotle's Fantasy is correct, then I differ from my past self in respect of some of my *time-indexed* properties; and so the improved or reinterpreted version of Leibniz's Law, restricted to time-indexed properties, still renders problematic my identity with my past self. We cannot, it seems, step twice into the same river unless it flows deterministically.

Before tackling this problem directly, we need to get straight about what Aristotle's Fantasy does and doesn't entail; so let's turn to that question. (Perhaps I should stop calling it Aristotle's "Fantasy," since I regard it as correct. But then again, Aristotle was Greek, and in Greek *phantasia* means what appears so, or is imagined to be so, without necessarily applying that it appears or is imagined falsely.)

Aristotle's Fantasy is often thought to contradict the law of Excluded Middle. No doubt some defenders of Aristotle's Fantasy have indeed thought themselves committed to rejecting Excluded Middle, but in my view this is too quick. Aristotle's Fantasy is most reasonably interpreted as maintaining that facts hold, or that propositions hold true (the difference need not detain us), *at times*, so that the complete specification of any assertion must make reference to the time of its holding. Let's introduce the dyadic predicate 'H' to illustrate this idea; 'H($p$, $t$)' will mean that proposition $p$ holds (is true, is the case) at time $t$. What Aristotle denies is the following claim, which I will, with some trepidation, call Bivalence:

**Bivalence:** $(\forall p)(\forall t)(\mathrm{H}(p, t) \vee \mathrm{H}(\sim p, t))$

Denying Bivalence, so defined, is by itself no violation of classical logic, since classical logic *per se* has nothing to say about this predicate 'H'. What classical logic insists on is the truth of Excluded Middle:

**Excluded Middle:** $(\forall p)(p \vee \sim p)$

But Bivalence, as I've defined it, is not a substitution instance of Excluded Middle. The following would be such a substitution instance:

$(\forall p)(\forall t)(\mathrm{H}(p, t) \vee \sim\mathrm{H}(p, t))$

But this latter claim entails Bivalence only if '~H($p$, $t$)' entails 'H(~$p$, $t$)' – and that, I take it, is precisely what Aristotle's Fantasy denies. Since 'H' is not a predicate of classical logic, it

follows that (my version of) Bivalence, which cannot be stated except in terms of 'H', is not a requirement of classical logic.

To make the difference between '~H($p$, $t$)' and 'H(~$p$, $t$)' more intuitive, think of 'H($p$, $t$)' as analogous to '$p$ is written in the Book of Fate.' From the fact that $p$ is *not* written in the Book of Fate, it does not follow that ~$p$ *is* written there; the page might simply be blank. Analogously, from the fact that a sea battle's occurring tomorrow does *not* hold now, it does not follow that a sea battle's *not* occurring tomorrow *does* hold now; tomorrow's page might still be blank, at least as far as sea battles go. (Presumably a sea battle's either-happening-or-not is already written on tomorrow's page, but '$(\forall p)(\forall t)(\mathrm{H}(p \vee \sim p, t))$' doesn't entail '$(\forall p)(\forall t)(\mathrm{H}(p, t) \vee \mathrm{H}(\sim p, t))$' either.)

Of course this analysis applies to propositions involving 'H' as well. Suppose $p$ is indeterminate at $t_1$, and does not acquire truth-value until $t_2$. Then the assertion 'H($p$, $t_2$)' will not hold at $t_1$, though it may come to hold at $t_2$. On this view, there is no position *outside* the temporal order, or at least none linguistically accessible to us, from which the truth of such a proposition could be assessed; our assertions are temporally bound all the way down. From an Aristotelean point of view, that's not a bug, it's a feature. (It's awkward for Thomists, of course, but that's their problem.)

Now we are in a position to consider how Aristotle's Fantasy and Leibniz's Law may be reconciled. Consider a time-indexed proposition, such as 'Fx at $t$'. In light of Aristotle's Fantasy, we must distinguish the time of what is being *stated* (the time of occurrence, as it were) from the time at which the proposition *holds*. If x's being F at $t_2$ is not yet a settled fact at $t_1$, then even after x becomes F at $t_2$, it will still be false to assert 'H(Fx at $t_2$, $t_1$)' – though of course true to assert 'H(Fx at $t_2$, $t_2$)'.

Not only, then, does x have, at $t_2$, a property (being F) that it lacks at $t_1$, x has, at $t_2$, a *time-indexed* property (being F at $t_2$) that it lacks at $t_1$. It thus runs afoul of even the improved version of Leibniz's Law. But Leibniz's Law can be improved again. Just as an unsophisticated reading of Leibniz's Law would rule out the acquisition of non-time-indexed properties, so an insufficiently sophisticated reading of Leibniz's Law would rule out the acquisition of time-indexed properties.

Consider Gauvain pacing before the tower at $t_1$, trying to decide whether to free Lantenac at $t_2$. Two possible futures lie open to him: one in which he has, at $t_2$, the property of freeing Lantenac at $t_2$, and one in which he has, at $t_2$, the property of declining to free

Lantenac at $t_2$. At $t_1$ he has, as yet, neither of those properties; but of course he does go on to free Lantenac (if you're in the middle of reading *Quatrevingt-treize* I apologise for the spoiler), and we want to say that the Gauvain who, at $t_2$, has the property of freeing-Lantenac-at-$t_2$ is numerically identical to the Gauvain who, at $t_1$, did not yet possess that property. And we *also* want to say, I think, that if Gauvain had chosen not to free Lantenac, then *that* Gauvain would have been identical with the earlier one. (Note that such a view is going to be inconsistent with a four-dimensionalist "spacetime worm" approach to identity over time. Here too, that's not a bug, that's a feature.) How can we restate Leibniz's Law so as to preserve its essential spirit while accommodating Gauvain's continuing identity?

Well, the point of Leibniz's Law is to rule out identicals having *inconsistent* properties. This desideratum appears to clash with Aristotle's Fantasy because lacking a property involves having the contradictory property. If x isn't F, then x must be ~F; since nothing can be simultaneously F and ~F, if x is F then nothing can be identical with x without likewise being F.

But since, as we've seen, '~H(*p*, *t*)' doesn't entail 'H(~*p*, *t*)', it follows that 'H(*p*, *t*)' and 'H(~*p*, *t*)' are not contradictories. What happens as time passes is not that Gauvain loses some time-indexed properties and gains others; whatever time-indexed properties he already has, he keeps *forever*. It's just that he acquires new ones too; as time passes he becomes more *determinate* than he was before. Just as Leibniz's Law had to be restated in terms of time-indexed properties in order to accommodate change in non-time-indexed properties, so it must now be restated in terms of *double-time-indexed* properties in order to accommodate change in time-indexed properties. By double-time-indexed properties I mean properties specified not only in terms of when a given property is possessed, but also in terms of when its possession *holds*.

Consider, then, the following two properties: a) freeing-Lantenac-at-$t_2$-holding-at-$t_1$, and b) freeing-Lantenac-at-$t_2$-holding-at-$t_2$. Gauvain lacks property (a) at $t_1$; in fact he has the contradictory property; call it (c). Property (c), of course, is not the property of *(not-freeing-Lantenac-at-$t_2$)-holding-at-$t_1$* – for then it would already be settled that Gauvain wasn't going to free Lantenac – but rather the property of *not-(freeing-Lantenac-at-$t_2$-holding-at-$t_1$)*. Property (c) Gauvain has not only at $t_1$ but *forever henceforth*. He will never lose property (c); he will never gain property (a). What he does acquire, if he frees Lantenac at $t_2$, is property (b). He acquires that property at $t_2$; he doesn't have it at $t_1$.

So at $t_1$ does he have the property contradictory to property (b)? Well, what *is* the property contradictory to property (b)? This is where things get tricky. At $t_2$, the property contradictory to freeing-Lantenac-at-$t_2$-holding-at-$t_2$ would appear to be not-(freeing-Lantenac-at-$t_2$-holding-at-$t_2$). But any property that Gauvain has at $t_1$ is going to have to be a property *holding at* $t_1$, and at $t_1$ it is *ex hypothesi* not yet settled what all of Gauvain's properties are going to be at $t_2$. To put it another way: if facts are temporally bound all the way down, as Aristotle's Fantasy seems best interpreted as maintaining, then any fact holding at $t_1$ must terminate, as it were, in a time-indexing to $t_1$. There can of course be facts, holding at $t_1$, *about* other facts holding at $t_2$, but a complete statement of the former would have to include a final time-indexing to $t_1$. For example, it may be a fact at $t_1$ that the sun's rising at $t_2$ will be a fact holding at $t_2$; but a full statement of what holds at $t_1$ would then have to specify not just the-sun's-rising-at-$t_2$-holding-at-$t_2$, but the-sun's-rising-at-$t_2$-holding-at-$t_2$-holding-at-$t_1$. All facts holding at $t_1$, whatever other times they may make reference to, must be embedded in a framework indexed to $t_1$.

Facts holding at $t_2$, by contrast, need not – indeed cannot – have $t_1$ as their *final* time-indexing. Hence there is no guarantee that a property possessed at $t_2$ will so much as *have* a property contradictory to it that could hold at $t_1$. On this understanding, the passage of time adds layers of time-indexing but doesn't remove any. Any fact holding at $t_1$ will still be a fact holding-at-$t_1$-holding-at-$t_2$ when $t_2$ rolls around; indeed, because that is the case this fact's holding-at-$t_1$-holding-at-$t_2$ already holds at $t_1$. But new facts can emerge at $t_2$ for which there is no negation assertible at $t_1$.

Recall property (b), which is the property of freeing-Lantenac-at-$t_2$-holding-at-$t_2$. *At* $t_2$, the property contradictory to property (b) would be property (d), the property of not-(freeing-Lantenac-at-$t_2$-holding-at-$t_2$). But since all facts are temporally embedded, if that property holds at $t_2$ then strictly speaking it's the property of not-(freeing-Lantenac-at-$t_2$-holding-at-$t_2$)-holding-at-$t_2$. In other words, its final time-indexing would be to $t_2$. Since there *are* no facts holding at $t_1$ whose final time-indexing is to $t_2$, there is just no such thing, at $t_1$, as having property (d). *Nothing could count.*

What there *is* such a thing as, at $t_1$, is property (e), which would be property (d) embedded in a final time-indexing to $t_1$. If we think of time as adding layers of time-indexing but not removing any, then whatever has property (e) at $t_1$ will acquire property (d) at $t_2$. But

property (e) is not the contradictory of property (b) either, and so Gauvain's acquiring (b) need not – cannot! – be a matter of his losing (e).

All Leibniz's Law needs to say, then, in order to avoid conflicting with Aristotle's Fantasy, is that A and B are identical only if, for any property one of them possesses, the other does not possess the *negation* of that property. (Leibniz's Law will of course be like all other facts in being temporally embedded; it too is not something timelessly assertible – though it will be assertible *at every time*.) Ordinarily not possessing the negation of a property means possessing the property; but if truth and meaning are temporally embedded as the Aristotelean position asserts, there will be cases where a property is so bound up with a later time that neither it *nor its negation* is available for possession at an earlier time.

My suggested solution to the conflict bears some analogy to Wittgenstein's conviction that apparently coherent questions like "is there a sequence of five consecutive sevens somewhere in the decimal expansion of π?" are in fact not coherent. For Wittgenstein, since mathematical truths are logically necessary truths, and the bounds of logic are the bounds of meaning, it follows that until we have established whether a mathematical statement is true we are not entitled to assume that it is so much as meaningful. Thus, contrary to appearances, the only way to give a coherent meaning to a mathematical conjecture is to prove it true. If we were to discover five consecutive sevens in the decimal expansion of π (or, as Wittgenstein would prefer to say, if we were to do something we were willing to call that), then the statement that this sequence of sevens exists would have meaning; but *from the standpoint of our present knowledge* the conjecture not only cannot be known, it cannot be so much as asserted. Similarly, then, if Aristotle's Fantasy is correct then the reason predictions of future contingents cannot yet be true or false is that they are propositions that can hold (and whose negations can likewise hold) only with a final time-indexing later than the present one. Nothing *counts*, at $t_1$, as the proposition that is verified or falsified at $t_2$.

But can't we talk meaningfully about future contingent events? Yes, certainly. But when we talk, at $t_1$, about what won't be causally settled until $t_2$, we're talking about facts holding-at-$t_2$-holding-at-$t_1$. On the Aristotelean view we can't yet talk about facts holding-at-$t_2$ *tout court*, because at $t_1$ such facts, with a final time-indexing to $t_2$, are simply not available to be referred to. Their only existence, at $t_1$, is in facts embedded in a final time-indexing to $t_1$. The passing of time brings with it the expansion of the semantic domain.