# The Incompatibility of Free Will and Naturalism

Jason Turner

*February 23, 2009*

This is a preprint of an article whose final and definitive form will be published in *The Australasian Journal of Philosophy* 2009; the Australasian Journal of Philosophy is available online at: http://www.tandf.co.uk/journals/.

**Abstract**

The *Consequence Argument* is a staple in the defence of *libertarianism*, the view that free will is incompatible with determinism and that humans have free will. It is often thought that libertarianism is consistent with a certain naturalistic view of the world — that is, even if libertarians are right, free will can be had without metaphysical commitments beyond those provided by our best (indeterministic) physics. In this paper, I argue that libertarians who endorse the Consequence Argument are forced to reject this naturalistic worldview. The Consequence Argument has a sister argument — I call it the *Supervenience Argument* — which cannot be rejected without threatening either the Consequence Argument or the naturalistic worldview in question.

## I The Consequence Argument

Determinism is the thesis that the laws of nature, when conjoined with any proposition accurately describing the state of the world at some instant, entail any other true proposition (cf. van Inwagen 1983: 58–65).[1] The Consequence Argument is supposed to show that this thesis is incompatible with free will. An informal version of the argument runs as follows:

> If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us (van Inwagen 1983: 56)

[1] Note that $p$'s entailing $q$ is here understood as the necessity of the conditional $p \rightarrow q$.

If the argument is sound, determinism is incompatible with free will.

This argument can be clothed in formal garb. This garb makes use of a modal operator, 'N', where 'N$p$' means '$p$, and no one has, or ever had, any choice about whether $p$' (van Inwagen 1983: 93).[2] And, following Alicia Finch and Ted A. Warfield (1998: 516), we can understand 'someone has a choice about whether $p$' as 'someone could have acted so as to ensure the falsity of $p$.'[3]

Originally, the Consequence Argument appealed to two inference rules 'N' was supposed to obey:

($\alpha$) From $\Box p$, deduce N$p$, and

($\beta$) From N$p$ and N$(p \rightarrow q)$, deduce N$q$

(van Inwagen 1983: 94), where '$\Box$' represents broad logical necessity. Unfortunately, these two rules are jointly invalid. Thomas McKay and David O. Johnson (McKay and Johnson 1996: 115) have shown that, given ($\alpha$) and ($\beta$), the N-operator is *agglomerative*:

(Agg) From N$p$ and N$q$, deduce N$(p \,\&\, q)$.

But (Agg) is invalid, as McKay and Johnson go on to show. In their example, we consider an agent who does not flip a coin but could have. In this case, N(*the coin does not land heads*) is true, and N(*the coin does not land tails*) is true. To ensure, for instance, the falsity of the coin does not land heads, one would have to ensure that the coin does land heads. This, presumably, is not something anyone could do. Yet N(*the coin does not land heads & the coin does not land tails*) is false. The agent could have ensured the falsity of the embedded conjunction by flipping the coin.[4]

In response, proponents of the Consequence Argument have rejected ($\beta$) in favour of another inference rule which does not entail (Agg):

[2]The arguments of this paper will be framed in terms of propositions rather than sentences; a more precise definition of 'N' would treat it as denoting a function from propositions to propositions where, if $p$ is a proposition and $\phi$ a sentence expressing $p$, N$p$ is the proposition expressed by $\ulcorner \phi$ and no one has or ever had a choice about whether $\phi \urcorner$. (Similar remarks apply for logical operations on propositions: $p \,\&\, q$, for instance, should be understood to refer to the proposition that conjoins $p$ and $q$.) However, I am going to be a bit sloppy about use and mention in the text, and will not continue with clarifications like this one in the notes. Such sloppiness reduces unnecessary clutter and is not uncommon in discussions of the Consequence Argument, as it does not affect the argument's content (cf. van Inwagen 2000: note 4; McKay and Johnson 1996: note 4).

[3]I take it that Finch and Warfield's talk of 'falsifying' a proposition is to be understood as 'ensuring the falsity of' a proposition.

[4]In fact, a similar counterexample can be used to directly show the invalidity of ($\beta$). When an agent doesn't flip a coin but could have, both of N(*the coin does not land heads*) and N(*the coin does not land heads* $\rightarrow$ *the coin is not flipped*) are true — the only way the agent could ensure the falsity

($\beta\Box$) From N$p$ and $\Box(p \rightarrow q)$, deduce N$q$

(Finch and Warfield 1998: 521–522; Widerker 1987: 41). If $P$ is a proposition that expresses the state of a world at a remotely early time (before there were any human agents, say), $L$ a conjunction of all the laws of nature, and $F$ any true proposition whatsoever, then the Consequence Argument is as follows:

**The Consequence Argument**

| | |
|---|---|
| (1) N$(P \,\&\, L)$ | Premise |
| (2) $\Box((P \,\&\, L) \rightarrow F)$ | Premise (Determinism) |
| (3) N$F$ | ($\beta\Box$: 1, 2) |

Thus, if determinism is true (and if no one has, or ever had, a choice about the truth of the conjunction of the laws of nature with a proposition expressing the state of the world in the remote past), then no one has ever had a choice about anything.

What of the first premise? It is highly intuitive that we cannot do anything to change the laws of nature — i.e., we cannot do anything that would ensure the falsity of the laws (and hence we 'have no choice' about them). It is likewise intuitive that we cannot do anything to change the past. Thus the truths of N$P$ and N$L$ are both intuitive.

Since, as we have already seen, the N-operator is not agglomerative, we cannot argue for the truth of N$(P \,\&\, L)$ by appealing to the truth of both N$P$ and N$L$. But while this is formally correct, it may not be much of an obstacle: (1) does not seem to be plausibly rejected even given the general invalidity of N-agglomeration.

Finch and Warfield (1998: 523) appeal to a 'broad past' principle in defence of (1): any proposition $p$ made true in the remote past, before anyone could have been around to act freely, is one for which N$p$ is true. Since $P$ & $L$ seems to be such a proposition (or so they claim), premise (1) is vindicated.

We might well wonder why we should accept this broad past principle (as well as the claim that $P$ & $L$ is part of the broad past). But Finch and Warfield do give an argument for its plausibility:

---

of the relevant conditional is by ensuring that the coin is flipped and lands tails. But N(*the coin is not flipped*) is clearly false. (Carlson 2000: 283–284; Crisp and Warfield 2000: 178–179). David Widerker (1987) gives an independent counterexample to ($\beta$), discussed by Timothy O'Connor (1993: 209) and McKay and Johnson (1996: 117–118).

> . . . it is important to be clear that the McKay and Johnson argument [against agglomeration] shows only that the inference from N$p$ and N$q$ to N$(p \,\&\, q)$ is invalid. This does not, by itself, provide any reason at all for thinking that N$P$ and N$L$ are true, while N$(P \,\&\, L)$ is not. An inspection of the difference [between the two cases] shows that the McKay/Johnson case seems to cast no doubt on the truth of N$(P \,\&\, L)$. In the McKay/Johnson case, one has no choice about either conjunct of a conjunction but does have control over the conjunction because although there is nothing one can do that would falsify either particular conjunct there is something one can do that might falsify either conjunct and would falsify the conjunction. . . [I]t is not at all plausible that though one cannot, for example, do anything that would falsify. . . the laws of nature, one might somehow do so. (1998: 523–524)

The idea, I take it, is that even though (Agg) is invalid, it is *defeasibly reliable*: if N$p$ and N$q$ are both true, then N$(p \,\&\, q)$ will be true too unless certain special conditions hold. More precisely:

> $(*)$ If N$p$ and N$q$ are both true and N$(p \,\&\, q)$ is not, then there is something that someone could have done which might have falsified $p$, might have falsified $q$, and would have falsified $p \,\&\, q$.

Then, if we already agree that N$P$ and N$L$ are true, the only way we can reject N$(P \,\&\, L)$ is by insisting that there is something that we could have done which might have falsified $P$, might have falsified $L$, and would have falsified $P \,\&\, L$. But the intuitions that we could have done nothing which *would* have falsified either $P$ or $L$ likewise tell us that nothing we could have done even *might* have falsified $P$ or $L$. From this and $(*)$ we can conclude N$(P \,\&\, L)$, and the argument goes through.[5]

[5]Retreat to ($\beta\Box$) is not the only way to revise the Consequence Argument in light of ($\beta$)'s invalidity: we might instead offer a different interpretation of 'N.' For example, McKay and Johnson (1996: 118-120) consider an interpretation on which 'N$p$' means '$p$ and no one could have acted in a way that *might* ensure the falsity of $p$' (see also Finch and Warfield 1998: 524–527; Huemer 2000: 538–540). And (van Inwagen 2000: §1) has proposed interpreting it as '$p$ and every region of logical space to which anyone has exact access is a subregion of $p$,' contrasting it with his gloss on the original reading: '$p$ and every region of logical space to which anyone has (not necessarily exact) access overlaps $p$.' Although I lack the space to argue for it here, anyone who finds the old argument convincing on any of these interpretations of 'N' ought to find Finch and Warfield's version convincing also (cf. Carlson 2000: 287).

# II Naturalism and the Supervenience Argument

## *A Naturalism*

In rough form, the thesis of naturalism holds that everything eventually boils down to fundamental physics. This is not necessarily a reductionistic view (although if you want to be a naturalist, being a global reductionist is one way to do it), but rather a thesis about what depends on what: all events and causal relations depend — and therefore supervene — on what happens at the fundamental physical level.

More precisely, call an event *microphysical* if it involves only the particles and properties postulated by fundamental microphysics — things such as electrons, quarks, bosons, etc., and their properties — and does not supervene on any other events that involve only these particles and properties.[6] Call a causal relation microphysical if it relates only microphysical events. In this case, naturalism (at least for our purposes here) comprizes:

(N1) All events supervene on microphysical events;

(N2) All causal relations supervene on microphysical causal relations; and

(N3) All (non-microphysical) events metaphysically depend on the microphysical events they supervene on.

Supervenience is understood as follows: if an event a supervenes on events $b_1, \ldots, b_n$, then it is impossible that $b_1, \ldots, b_n$ occur and a fail to occur.[7] Unfortunately, 'metaphysical dependence' is not so easily defined, although the idea is not completely obscure. We have an intuitive grasp on the thought that wholes somehow depend on their parts for their existence, sets somehow depend on their members for their existence, and so on. The idea behind (N3) is that the relationship between an event and its supervenience base is relevantly like that thought to hold between a set and its members or a composite object and its parts. Just as, if a whole or a set exists, it does so thanks to the existence of

[6] This second clause essentially guarantees that these events are 'minimal': even if my raising my right hand is ultimately comprized of tiny particles changing their fundamental physical properties, since its being so comprized entails its supervening on the events involving those particles and properties, it does not count as microphysical.

[7] As long as events are closed under (the event counterpart) of negation — that is, as long as for every (possible) event $e$ there is also the (possible) event of $e$'s non-occurrence (compare the discussion of omissions and thesis (T1) below) — this claim follows from standard formulations of supervenience (see Bricker 2006: 267–270; cf. Kim 1984*a*: 64; Kim 1987: 81–82).

its parts or its members (and not *vice versa*), if a macrophysical event occurs, it does so thanks to the occurrence of the events in its microphysical supervenience base, and not *vice versa*.

There are stronger and weaker versions of each of naturalism's claim. The strong form of (N1) holds that every event *metaphysically* supervenes on the microphysical — that is, any two metaphysically possible worlds containing different events contain different microphysical events as well. The weak form requires only nomic supervenience between events, so that any two possible worlds with divergent events have either divergent microphysical events or divergent laws of nature.[8] Likewise, a strong version of (N3) holds that the metaphysical dependence of the macrophysical on the microphysical is independent of the laws, whereas the weaker form allows that the dependence may be somehow nomic or law-relativized.

Claim (N2) also comes in a variety of strengths, the strongest holding that for any events $c$ and $e$, $c$ caused $e$ if and only if $c$'s microphysical supervenience base caused $e$'s microphysical supervenience base ([cf. Kim 1984*b*: 262). A weaker version holds merely that if $c$ causes $e$, then there is a microphysical causal chain running from some events in $c$'s supervenience base to some events in $e$'s supervenience base which forms (a perhaps, but not necessarily, proper) part of a complete microphysical causal chain for $e$'s supervenience base.

Call the version of naturalism generated by combining the weaker versions of these three theses weak naturalism. Weak naturalism is weaker than it might be, but stronger than some other views that have gone under the title of 'naturalism.' Of especial interest to us here are the self-proclaimed 'naturalistic' agent-causal theories of Timothy O'Connor (2000) and Randolph Clarke (2003). Insofar as both of these theories are consistent with something like (N1), there is a sense in which they deserve to be called 'naturalistic.' Neither, however, is (or claims to be) to be consistent with (N2) or (N3), since they both posit forms of substance causation not meant to supervene on microphysical event-causation.[9] As I use 'naturalism' here, I mean to rule agent-causal theories, including O'Connor's

[8] I leave it open how large a given macro-event's supervenience base must be, but I take it that any sensible naturalist will want its microphysical supervenience base to be smaller than the totality of micro-events in a world (cf. Stalnaker 1996: 228–230).

[9] One could posit a form of irreducible agent-causation which nonetheless supervenes on an event-causal microphysical base. (Any agent-caused event would in this case be, in some sense or other, overdetermined.) Although (N2) does not rule out this sort of agent-causal account, I do not intend the Supervenience Argument to be effective against it. It may well be that (N3) does manage to rule this out; if not, (N2) should be augmented with the claim 'and all causation is event-causation.' Thanks here to Tom Crisp.

and Clarke's, out. The Supervenience Argument is supposed to undercut libertarian theories that attempt to secure free will without adding metaphysical commitments that go beyond those of current science. It is not intended to target agent-causal or other theories that 'postulate unusual forms of... causation' (Kane 1996: 115), no matter how friendly to current science those theories may otherwise try to be.

## *B The (Informal) Supervenience Argument*

The Supervenience Argument is designed to show that, if weak naturalism is true, the class of actions about which someone has or ever had a choice is empty. Since stronger forms of naturalism entail the weak form, it follows that any form of naturalism precludes actions of this sort. As with the Consequence Argument, there is an informal version of the Supervenience Argument:

> If weak naturalism is true, then our acts are the consequences of the laws of nature, events in the remote past, and the outcomes of undetermined microphysical events. But it is not up to us what went on before we were born, what the laws of nature are, or how undetermined microphysical events turn out. Therefore, the consequences of these things (including our present acts) are not up to us.

This argument, or at least something very much like it, has already found favour in the eyes of some. Trenton Merricks, for instance, gives what I take to be a version of it in *Objects and Persons* (Merricks 2001: 155–161). Others have given related arguments against the compatibility of free will and a broadly naturalistic worldview (given the incompatibility of it with determinism) (see e.g. Bishop 2003, Loewer 1996, Unger 2002).

Relatedly, J. A. Cover and John Hawthorne (1996) have argued that the combination of something like (N1) and agent-causation is coherent but implausible. There are many points of contact with the argument I give below, but their argument for the implausibility of the combination hinges crucially on the interaction of agent-causation with various positions in the philosophy of mind (see esp. 62–70) — issues I hope to remain neutral on here.[10]

[10]Their are other crucial differences between Cover and Hawthorne's argument and mine. For instance, they have as a premiss that whether or not an agent acts freely supervenes on the intrinsic microphysical history of that agent (59). The argument to be given below need not take a stand on whether or not the microphysical supervenience base in question be intrinsic to the agent; see note 8 above.

None of these authors, however, has taken the trouble to clothe the Supervenience Argument in the same formal garb the Consequence Argument wears.[11] As a result, we do not yet have a clear idea of how closely the Supervenience Argument is tied to the Consequence Argument and whether or not there is some move to be made that will undermine the former while leaving the latter intact. In the balance of this paper I provide this formal clothing. But we need to clarify a couple of issues before the argument will be ready for its new outfit.

## C *Choosy Actions*

Call an event a choosy if and only if $A$ & $\sim$N$A$ is true, where $A$ is a proposition expressing the occurrence of $a$.[12] The Supervenience Argument is going to go like this: If there are any choosy actions, there is a first one. But if naturalism is true, the first choosy act is the consequence of things that went on before it and certain undetermined microphysical events, neither of which anyone ever had a choice about. Thus, if naturalism is true, there are no choosy acts and people are not free.

This argument thus depends for its soundness on the following theses:

(T1) If anyone is free, there are some choosy actions; and

(T2) If there are some choosy actions, there is a first one.

I take it that (T1), or something very close to it, is required if the Consequence Argument is supposed to be about free will. The Consequence Argument is supposed to show that there are no choosy propositions if determinism is true; but surely this is primarily important because some propositions are about what we do, and so the lack of choosy propositions entails the lack of choosy actions.

Granted, there are some slight subtleties here: perhaps omissions aren't actions, and perhaps in some possible world people make all sorts of choosy omissions (that is, they omit performing acts $a$ where *a didn't occur & $\sim$N(a didn't occur)* is true) but never, in fact, perform any choosy action. But it should be clear

[11] Cover and Hawthorne do discuss ($\beta$) in their argument (58–63), although it is not immediately obvious how to revise their discussion in light of ($\beta$)'s invalidity. It is also unclear how closely tied the principles applied in the latter parts of their argument are to the Consequence Argument.

[12] As an anonymous referee pointed out, $A$ had better not express $a$'s occurrence under certain sorts of descriptions. If my raising my right arm is the first bid at the auction, then I may have been able to ensure the non-occurrence of that arm-raising without being able to ensure the falsity of the proposition that *the first bid at the auction occurred*. So let $A$ be singular with respect to $a$: that is, let $A$ be the proposition expressed by 'x occurred' when 'x' is assigned to $a$.

that these sorts of subtleties can be ironed out (by replacing 'actions' for 'actions and omissions', for instance); since omissions are both caused by and causes of other things (Schaffer 2000), making this modification throughout won't affect the argument. But once we see that we can make the modification if needed, we see we needn't bother; better to simply assume, to reduce clutter if nothing else, that omissions are a sort of action.

(T2) also looks plausible, modulo similar subtleties. For instance, on the so-called 'fine-grained' theory of action, every time anyone does *anything*, they do infinitely many things simultaneously: e.g., they raise their arm, they raise their right arm, they raise their right arm in a room with a temperature of over 32°C, etc., all of which count as separate actions according to the fine-grained theorist. But again, a simple revision cures all. Where a coarse-grained theorist sees a single action, the fine-grained theorist sees many that are all connected in a certain way (all on the same 'action-tree' (Ginet 1990: 19, 46–52)): instead of talking of the first choosy action, we can talk of the first choosy equivalence class of actions thus connected. Again, once we see what to do to fix for the eventuality, we see we do better to assume the coarse-grained theory for simplicity sake.

But mightn't there be genuine ties for the first choosy action? Sure; but if so, they will be causally independent of each other and so we can arbitrarily pick one as 'first' for the sake of argument without doing any damage.

One final point: a clever philosopher might suggest that (T2) is false because people have been around forever performing choosy actions. Then there wouldn't be a first choosy action. Or maybe she will say there was once some extremely fast mover who, at every point before and up to time $t$ had never performed a choosy action, but at every time after $t$ had performed some choosy actions. If time is dense — that is, if there is no 'first' time after $t$ — then again there wouldn't be a first choosy action.

Since these hypotheses are empirically falsified (we haven't been around that long[13] and just can't move that fast), these clever suggestions don't impugn the *truth* of (T2). Since they do seem possible, though, they may impugn the *necessity* of (T2). If so, they expose a certain weakness in the argument: it won't show

[13] A minor technicality: current scientific knowledge may leave it open that, in the actual world, time really has gone back forever and at any time, there is some previous time at which there were agents going around and doing things. The Big Bang could have been preceded by a Big Crunch, which was itself preceded by a previous Big Bang, and so on back forever. And every cycle might have included people. Still, since Big Crunches and subsequent Big Bangs wipe out relevant causal connections between the people who came before and those that came after, it won't be hard to revise either the Consequence or the Supervenience Argument in a way that makes this eventuality irrelevant.

the impossibility of the conjunction of naturalism with the claim that people act freely, but rather the impossibility of the conjunction of naturalism with the claim that *people like us*, in a *universe like ours*, act freely.

But this weakness isn't distinctive of the argument to be given; the Consequence Argument already exhibits it (Warfield 2000). For example, if time went forever backwards, and if there were no time before which there were people acting, a friend of the Consequence Argument would have little justification for claiming N*P* (Campbell 2007). Nonetheless, the Consequence Argument still seems to show something — namely, that people like us in a universe like ours (i.e., a universe with a time before which there were no people) can't be free if determinism or naturalism, respectively, is true. Since we are like us and we live in a universe like ours, I think (pace Warfield's (2000) claim to the contrary) we ought to be concerned about these claims and therefore concerned about both of these arguments. And this means we can rely on the truth of (T2) without worrying about its possible falsity in worlds quite different from ours.

## *D Causal Relations*

By (T2), if there are some choosy actions, there is a first one. Call it *r*, and suppose it was performed by an agent *S*. For illustration, suppose that the causal theory of action is true.[14] Then *r*, by virtue of being an action, will have been caused by some particular pair of desires and beliefs (or their neural realizers), which I will call *db*. But *db* probably will not encompass all of the causes of *r*. Causal theorists seldom think that a desire/belief pair alone is nomically sufficient for an action. Other inner states of the agent, as well as external, environmental factors, etc., will likely figure into the causal story of most actions. So let *db+* represent the sum total of what we would call the causes of *r* if we knew enough about *r*'s production.

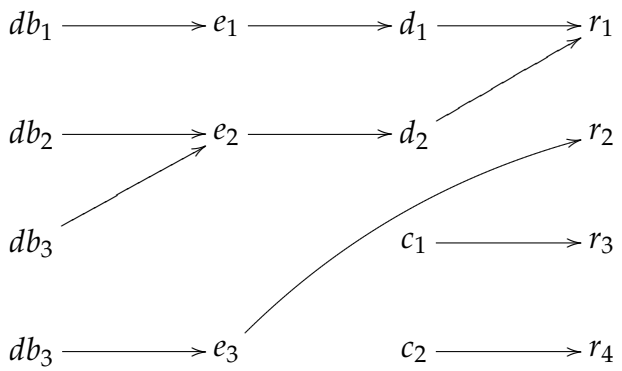
If weak naturalism is true, both *db+* and *r* will supervene on microphysical events. Any realistic discussion of *db+*'s and *r*'s supervenience bases, and the causal connections between them, are going to be hopelessly complicated. So let's just make up a simple story to be getting on with. When it comes to these

[14]This supposition is just for illustration: many non-causal theories of action (e.g. Ginet 1990) are consistent with actions *having causes*, even though *being an action* does not depend on having any particular type of causal history. I do assume, however, that every event (actions included) has some cause or another; I did not include this assumption in the statement of naturalism, but they seem to naturally go together. Given this assumption, even if we reject the causal theory of action we can take *db+* to be whatever the causes of *r* in fact are — without effecting the argument whatsoever.

events' supervenience bases, let's suppose they each have only four events in them ($db_1, \ldots, db_4$ and $r_1, \ldots, r_4$ respectively).

Since $db+$ caused $r$ and causal relations supervene on microphysical causal relations, there will be some microphysical causal chain running between $db+$'s supervenience base and $r$s supervenience base. Of course, not everything in $r$'s supervenience base need have been caused by something in $db+$s supervenience base; let us suppose that $r_1$ and $r_2$ are caused by $db_1, \ldots, db_4$, whereas $r_3$ and $r_4$ are not. And suppose that the causal chains work like this: $db_1$ causes an event, $e_1$, which in turn causes another event, $d_2$. Meanwhile, $db_2$ and $db_3$ jointly cause an event $e_2$, and $db_4$ causes an event $e_3$. Then $e_2$ causes $d_2$ and $e_3$ causes $r_2$. Finally, $d_1$ and $d_2$ jointly cause $r_1$. Suppose further that $c_1$ and $c_2$ are the causes of $r_3$ and $r_4$, respectively. The causal chain between $db+$'s supervenience base and $r$'s supervenience base will then be as depicted in figure 1.

Figure 1:



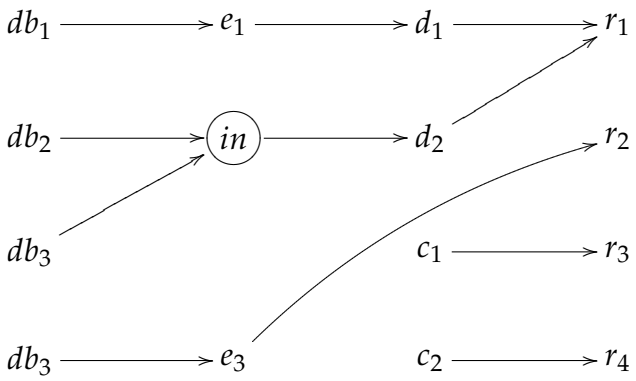

Now, if this is a deterministic universe, the occurrence of $db_1, \ldots, db4, c_1$, and $c2$ will be nomically sufficient for the occurrence of $r$'s supervenience base (and therefore nomically sufficient for $r$). If we are talking about choosy events, however, libertarians will hasten to remind us that the universe — and this causal chain in particular — had better not be deterministic. So we shall suppose it is not.

If $r$ is not going to be deterministically caused, then some link in the microphysical causal chain will have to be indeterministic. There are two places indeterminism could crop up. First of all, one of the events in the causal chain 'in between' $db+$ and $r$ may have been only indeterministically caused by its

antecedents. Or the indeterminism could occur 'at the end' of the chain: one of the events in $r$'s supervenience base may have been only indeterministically caused.

Let us begin by supposing that the indeterminism is of the first kind — the kind that only crops up 'in the middle.' Suppose, for example, that $e_2$ is the indeterministic culprit, and call it in from here on out to emphasize its indeterministic nature.[15]

Figure 2:




Now the causal chain looks as depicted in figure 2: events $db_2$ and $db_3$ cause $in$, but only indeterministically. There are worlds with the same laws of nature in which $db_2$ and $db_3$ occur but in does not. In this case, the occurrence of the $db$s and the $c$s will not be nomically sufficient for the occurrence of the $r$s. However, the occurrence of the $db$s, the $c$s, and *in will* be sufficient for the occurrence of the $r$s. Therefore, if $DB$ is a proposition expressing the occurrence of $db_1, \ldots, db_4$, $C$ a proposition expressing the occurrence of $c_1$ and $c_2$, $IN$ a proposition expressing the occurrence of $in$, and $R_1, \ldots, R_4$ propositions expressing the occurrence of each of $r_1, \ldots, r_4$, respectively, then the following proposition is true:

$$\Box((DB \;\&\; C \;\&\; IN \;\&\; L) \rightarrow (L \;\&\; R_1 \;\&\; R_2 \;\&\; R_3 \;\&\; R_4))$$

[15] I assume a simple model of indeterminism which essentially throws a 'Democritan swerve' into classical particle mechanics. As far as I can see, for present purposes this is a perfectly adequate simplification of the way things would work on, for instance, the truly indeterministic interpretation of quantum mechanics proposed by (Ghirardi et al. 1986). I doubt that any other interpretation of quantum mechanics will provide the libertarian with a substantially different yet fully naturalistic (in the sense of (N1)–(N3) above) implementation of indeterminism. I cannot argue for that here, but hope to do so in future work (but see also Loewer 1996).

Likewise, since $r$ supervenes nomically on $r_1, \ldots, r_4$,

$$\Box((L \,\&\, R_1 \,\&\, R_2 \,\&\, R_3 \,\&\, R_4) \rightarrow R)$$

is also true. And these two propositions together imply

$$\Box((DB \,\&\, C \,\&\, IN \,\&\, L) \rightarrow R).$$

## *E The (Formal) Supervenience Argument: A First Pass*

We are now in a position to offer a formal version of the Supervenience Argument. The first premiss is the nomic sufficiency of the *db*s, the *c*s, and *in* for $r$ defended above. The second is that no one can, or ever could have, ensured the falsity of $DB \,\&\, C \,\&\, IN \,\&\, L$. Since this proposition is made true before the first choosy act, it falls within the 'broad past' and so nobody ever had a choice about it. Thus, $\mathrm{N}(DB \,\&\, C \,\&\, IN \,\&\, L)$ is true. The argument now consists of a single application of ($\beta\Box$):

**The Supervenience Argument**

| | |
|---|---|
| (1) $\Box((DB \,\&\, C \,\&\, IN \,\&\, L) \rightarrow R)$ | Premiss (Naturalism) |
| (2) $\mathrm{N}(DB \,\&\, C \,\&\, IN \,\&\, L)$ | Premise |
| (3) $\mathrm{N}R$ | ($\beta\Box$: 1, 2) |

So $r$ is not a choosy act. Recall, though, that r was supposed to be the first choosy act; it follows (by (T2)) that there are no choosy acts at all. If the argument is sound, no one has, or ever had, a choice about anything.

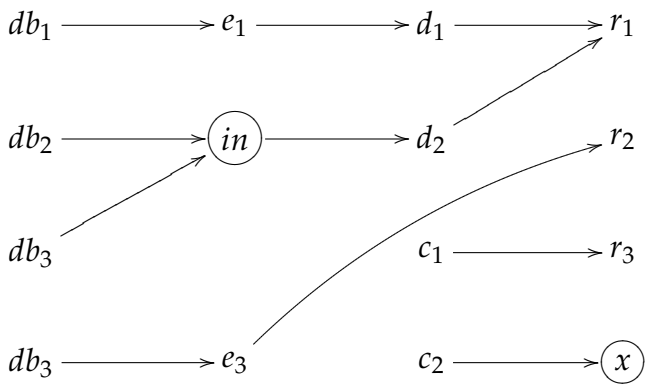
Someone may object that $DB \,\&\, C \,\&\, IN \,\&\, L$ is not part of the *remote* past, since it occurs very soon before $r$, and that therefore the broad past principle doesnÕt license premiss (2). This appeal to the remoteness of the past is a red herring. It is not as though we think the recent past is only somewhat fixed, and we can change it a bit, whereas as time goes on it 'solidifies' until it is eventually unchangeable. Rather, *the past* — remote or not — cannot be changed by *anything we can do now*. The only reason to appeal to a 'remote' past in defence of the Consequence Argument is to make sure that we do not appeal to a time at which people (not necessarily we) were going around performing choosy actions. If our proposition is made true before the first choosy action, though, we are in the clear.

## F The Argument for Trickier Cases

In the above argument, I supposed there was a microphysical causal chain between the *db*s and the *r*s with indeterministic links only 'in the middle.' But there may instead be a causal chain from the dbs up to but not including the *r*s in which one of the *r*s *itself* is the undetermined link. This means that there could be two possible worlds (with the same laws of nature) in which the entire causal chain strictly between the *db*s and the *r*s occurred, but (all of) the *r*s occur in only one of them. Perhaps at least one of the *r*s is not determined by its causes.

Suppose $r_4$ is the undetermined event. (We shall call it $x$ from here on, thus making $r$'s supervenience base $r_1$, $r_2$, $r_3$, and $x$, as shown in figure 3.) Suppose

Figure 3:




for now that nobody can, or ever could have, ensured the falsity of $X$, the proposition that expresses the occurrence of $x$. (We will return to this premiss in a moment.) What I would like to do is agglomerate $\mathrm{N}(DB \,\&\, C \,\&\, IN \,\&\, L)$ and $\mathrm{N}X$, which would allow me to offer the following argument:

**The Tricky Supervenience Argument**

(1) $\Box((DB \,\&\, C \,\&\, IN \,\&\, L \,\&\, X) \rightarrow (L \,\&\, R_1 \,\&\, R_2 \,\&\, R_3 \,\&\, X))$ Premiss (N2)

(2) $\mathrm{N}(DB \,\&\, C \,\&\, IN \,\&\, L \,\&\, X)$ Premiss

(3) $\mathrm{N}(L \,\&\, R_1 \,\&\, R_2 \,\&\, R_3 \,\&\, X)$ ($\beta\Box$: 1, 2)

(4) $\Box((L \,\&\, R_1 \,\&\, R_2 \,\&\, R_3 \,\&\, X) \rightarrow R)$ Premiss (N1)

(5) N$R$ ($\beta\Box$: 3, 4)

The first premiss is unproblematic: $DB$ & $C$ & $IN$ & $L$ entails $R_1$ & $R_2$ & $R_3$ since $r_1$, $r_2$, and $r_3$ are nomically necessitated by the causal chain leading up to them, and $L$ & $X$ trivially entails $L$ & $X$. The premiss in the fourth line of the argument is equally unimpeachable, since it simply expresses the nomic supervenience of $r$ on $r_1$, $r_2$, $r_3$, and $x$.

What of the second premiss? Of course, I cannot straightforwardly appeal to (Agg) to get it from N($DB$ & $C$ & $IN$ & $L$) and N$X$. But consider ($*$) from above: according to it, I can go ahead and agglomerate these unless there is some action $a$ that (i) $S$ could have performed, (ii) if $S$ had performed it, it might have ensured the falsity of $DB$ & $C$ & $IN$ & $L$, (iii) if $S$ had performed it, it might have ensured the falsity of $X$, and (iv) as a result, if $S$ had performed it, it *would* have ensured the falsity of their conjunction. But there can be no such action $a$. Condition (ii) says that it would have to be an action that *might* ensure the falsity of $DB$ & $C$ & $IN$ & $L$. But this proposition lies in the broad past and therefore nothing anyone can do even might ensure its falsity. So ($*$) will let me agglomerate anyway; (2) is vindicated and the argument follows.

Some may object: since *in* and $x$ are both undetermined, perhaps some action $a$ that $S$ performed *before* $r$, and therefore could have performed, might have ensured the falsity of $IN$ and might have ensured the falsity of $X$, and it was only bad indeterministic luck that $a$ did not ensure that one of them was false. But even if this is so, it does not undermine the appeal to ($*$). The mere fact that there is something someone could have done which might have ensured the falsity of $p$ and might have ensured the falsity of $q$ does not show that, if they did it, it would have ensured the falsity of $p$ & $q$. It's not enough to meet conditions (i)–(iii); (iv) must be met, too. Suppose Herbert does not toss a six-sided die but could have. Then he could have done something — toss the die — which might have ensured the falsity of *the die does not land one* and which might have ensured the falsity of *the die does not land six*. But it's just wrong to say that his toss would have ensured the falsity of *the die does not land one & the die does not land six* — the die might have landed four instead, in which case both conjuncts would still have been true. To see that the case under consideration is more like a die-tossing than a coin-tossing case, note merely that $a$, whatever it might have been, was performed but $DB$ & $C$ & $IN$ & $L$ & $X$ was true. So it's just not the case that if someone had performed a it would have ensured the falsity of this proposition.

## G The Status of NX

The argument above depends on the truth of N$X$ — the truth of the claim that nobody could have done anything that would have ensured the falsity of $X$, and therefore (we may suppose) that nobody could have done anything that would have ensured the non-occurrence of $x$. Why should we buy this claim?

Let's start in a (not so obvious) place: by asking why we should accept ($\beta\Box$). I take it that the intuitive idea behind ($\beta\Box$) goes something like this: 'Look, suppose you were able to ensure the falsity of *the match catches fire at 2 p.m.* There are a whole bunch of facts that, all together, made the case that the match catches fire at 2 p.m.: that it wasn't water-soaked, that there was oxygen in the room, that it was struck at 2 p.m., that the laws of nature have it that matches struck in those circumstances catch fire, etc. So the only way you could have ensured the falsity of *the match catches fire at 2 p.m.* was if you had ensured the falsity of *the match wasn't water-soaked at 2 p.m.*, or of *there was oxygen in the room at 2 p.m.*, or so on. If you are to be able to ensure the falsity of some proposition $p$, you have to be able to ensure the falsity of one of the many facts that come together to make $p$ the case.'

This is not to say, as McKay and Johnson's coin-tossing example reminds us, that there must be some *particular* case-making fact or another you must be able to ensure the falsity of. Perhaps you can do something that will ensure the falsity of *the match wasn't water-soaked at 2 p.m. and there was oxygen in the room at 2 p.m.* even though it's not the case that it would have falsified either of these conjuncts (because it might have falsified one or might have falsified the other). But if nothing you could do even had a hope of falsifying at least one of the relevant facts, nothing you could do would even have a hope of ensuring the falsity of *the match catches fire at 2 p.m.*

If determinism is false, then although some facts might be *part* of what makes it the case that $p$, there may not be any *complete* set of facts that make $p$ the case. For instance, part of what makes it the case that a certain particle decays at a time $t$ might be that the particle exists at that time. That's why I can ensure the falsity of *the particle decays at t* — I can destroy the particle before $t$ arrives. But nothing makes it *entirely* the case that the particle decays at $t$; that's part of why we think it is undetermined.

A canny libertarian, of course, will think that we can sometimes ensure the falsity of a proposition $p$ even though nothing we can do even might ensure the falsity of any of the things that (partially) make it the case that $p$. Perhaps my current mental states constitute a good chunk of what makes it the case that

I walk on past the hungry beggar without offering assistance. Nonetheless, if my action is choosy, it may be that I could have stopped instead even though nothing I could do even might have made it the case that I had different mental states before stopping. In fact, the libertarian ought to think that I could have helped the beggar *even though* for every $p$ that partly made it the case that I didn't help, nothing I could do even might have ensured the falsity of $p$. In a situation like this, let's say that I have *direct control* over the proposition *I walk on past the beggar*.[16] If I have direct control over a proposition, I can ensure its falsity without being able to do anything that even might ensure the falsity of anything else that is part of what makes it the case. However, if I don't have direct control over a proposition, then in order to ensure its falsity, I must be able to do something that at least might ensure the falsity of one of the things that makes it be the case.

I take it that thoughts like these, which can be captured with the principle:

> (B) If someone is able to ensure the falsity of $q$, then either she has direct control over $q$, or she is able to do something that might ensure the falsity of some $p$ where $p$ is (part of) what makes it the case that $q$,

are primarily what lie behind ($\beta\Box$)'s intuitive appeal. They are also what underwrite the intuitive appeal of a principle such as ($*$). Why think that just because we can ensure the falsity of some conjunction, we must thereby be in a position to do something that at least might ensure the falsity of one of its conjuncts? Plausibly, because we think a conjunction is *made true* by its conjuncts, so if you are going to do something that ensures the falsity of a conjunction, the thing you do must ensure the falsity of at least one of the conjuncts (even if you have no control over which conjunct will be the falsified one).[17]

---

[16] More plausibly, I have direct control over the proposition *I decide to walk on past the beggar*; presumably part of what makes my overt actions the case are the prior decisions I make, and it is these I ultimately have direct control over. More on this below.

[17] A referee suggested that (B) might not provide the intuitive underpinning for ($\beta\Box$) on the grounds that entailments often run 'in the opposite direction' of the making-it-the-case relation. Note, however, that this does not bar (B) from capturing the intuitions driving the libertarian's acceptance of ($*$). Furthermore, ($*$), plus the plausible principles

> (i) Necessarily, if an action a ensures the falsity of $p$, then $p$ is false,
>
> (ii) Necessarily, if an action $a$ ensures the falsity of $p$, and if $p$ is a truth-functional consequence of $q$, then $a$ also ensures the falsity of $q$,

can together be shown to entail ($\beta\Box$). It seems that (B) provides an intuitive underpinning for ($*$) and ($*$) provides part of a *logical* underpinning for ($\beta\Box$).

If (B) is correct, then N$X$ is highly implausible. For if N$X$ is false, someone is able to ensure the falsity of $X$. In this case, by (B), either she has direct control over $X$, or she is able to do something that might ensure the falsity of some proposition that is part of what makes it the case that $X$. Since nothing she can do even might ensure the falsity of any of the propositions that are part of what make $X$ the case (in this case, propositions about $x$'s causal history, which are in $x$'s past), she must have direct control over $X$. But it is not at all plausible that we have direct control over microphysical events, so it is not at all plausible that N$X$ is false.

What is so implausible about having direct control over microphysical events? When someone has direct control over a proposition $p$, they can make sure that $p$ is false without ensuring the falsity of some $q$ that is part of what makes $p$ the case. By contrast, when someone has non-direct control over $p$, they can ensure the falsity of $p$, but only by ensuring the falsity of some $q$ that is part of what makes $p$ the case. As a result, when an agent $S$ has only indirect control over $p$, there is an informative answer to the question, 'How would $S$ make it the case that not-$p$?' If you ask me 'How would you make it the case that the match does not light?', I can intelligibly respond with 'I would remove the oxygen from the room' or 'I would soak the match in water' or so on. By contrast, when $S$ has *direct* control over $p$, there is no informative answer to the question, 'How would $S$ make it the case that not-$p$?'

In general, if someone claims to be able to ensure that not-$p$, it seems appropriate to ask how she would make not-$p$ be the case. There is a small but notable class of exceptions: propositions that express the occurrence of certain kinds of mental events, such as willings, decidings, and intendings. Suppose you claim you could have made it the case that you decided to pass the beggar by (instead of helping, as you in fact did). I ask you how you would have made this the case. 'By deciding to pass by', you would respond. It would be plainly inappropriate for me to ask for any further explanation. Plausibly, if you have control over anything, you have direct control over your decisions and the like: there is nothing *prima facie* mysterious about the thought that you have control over mental actions such as decisions which doesn't depend on control over other propositions that make those decisions occur. If I ask you how you can decide to pass on by, you are perfectly within your rights to simply respond that you do it by deciding — nothing further is required.

By contrast, if you claim to be able to keep an electron from gaining a certain spin property (for instance), you do seem to owe the rest of us an explanation of how you would do this. Perhaps you could, as in Widerker's (1987) example, do

it by destroying the electron ahead of time. But if, when challenged, you respond that you can keep the electron from gaining this property just by keeping it from gaining the property — well, this would be a very mysterious power indeed. If you have direct control over microphysical events, though, then this is just the sort of power you have.

It is incredibly implausible that we have such direct control over microphysical events; as such, we ought to accept N*X*, and the argument follows. But I commonly hear two sorts of objections to this claim. Let's tackle them in turn.

### Objection 1

Some object that direct control over microphysical events is not so mysterious after all, on the grounds that mental actions such as decidings and willings might be identifiable with, say, the firing of a single neuron. If such a firing is itself a microphysical event, and if there is nothing mysterious in having direct control over decidings and willings, then there is nothing mysterious about having direct control over microphysical events.

It is doubtful in the first instance that we can plausibly identify these mental events with single neuron-firings: there is good reason to think that even the simplest cognitive activities require a few hundred neuronal stimulations at a minimum.[18] But, more importantly, even if such a mental event could be identified with the firing of a neuron, that firing is still a complex electrochemical event involving, and supervening on, events involving very many subatomic particles. Neuronal events thus do not count as 'microphysical' on the definition given in section II.A.

Could mental actions such as decidings or willings be identified with *bona fide* microphysical events, such as the change in energy of a single electron? Given the difficulties identifying such actions even with single neuron firings, it seems unlikely. But there are deeper reasons to avoid such an identification. Neurons comprize the smallest computational units of the brain; as such, neuronal events ought to comprize the smallest computational brain events. But it would be strained indeed to identify a mental action with an event smaller than the smallest computational unit in the brain. Since neurons are the smallest computational units in the brain, and microphysical events are smaller, it is very

[18]See, e.g., Huber et al. (2007). Huber and colleagues did report achieving their results with the stimulation of far fewer neurons (about sixty or so) when the stimulation was over a longer period of time. But the lengthened time suggests that there was far more firings per neuron and thus does not help the suggestion that a mental action could plausibly be identified with a single neuronal event.

difficult to think of microphysical events as actions, even of the mental kind.

This, of course, is no argument against the *possibility* of some microphysical events being actions. Perhaps there could be possible agents with brains that use subatomic particles as their basic computational units, in such a way that some mental actions could also be microphysical events. But if these possible agents are very different from us, this would give us no reason to think that *we* could be free in a naturalistic world — which, as I argued in section II.C, is the compatibility claim we should be interested in. Insofar as our best science suggests such possible agents are very different from us, this gives us reason to think that a microphysical event can be no action of ours and hence that we cannot have direct control over $X$. But if this is right, then since we also cannot ensure the falsity of anything that makes $X$ partly the case, we ought to conclude that N$X$ is true.

### Objection 2

A second objection insists that it is no mystery that we have direct control over microphysical events: we have it by having control over the macro-events that they supervene on. That is, if you ask how $S$ is able to make it the case that $X$ is false, the answer is that $S$ does it by refraining from $r$-ing. That is, if $S$ refrains from $r$-ing, this will ensure the falsity of $X$, because it will ensure the non-occurrence of $x$. $S$ can keep $x$ from occurring by refraining from $r$-ing — no mystery here.

Granted, we often think that I can exercise control over the microphysical by doing various macrophysical things. I can, for instance, make certain microphysical events happen in my nerves by wiggling my toes, make complex microphysical events happen in my brain by thinking about daffodils, and so on.

But for this observation to undermine the above considerations, it must be the case that, if someone ensures the non-occurrence of some microphysical event by performing a macrophysical event, then the occurrence of the macrophysical event is part of what makes it the case that the microphysical event occurred. It is entirely consistent, though, with $S$'s (potentially) ensuring the falsity of $x$ by refraining from $r$-ing that the occurrence (or non-occurrence) of $x$ is part of what makes it the case that $r$ occurs (or does not occur). That is, '$S$ does $a$ by $b$-ing' need not amount to the 'the occurrence of $b$ is part of what makes it the case that $a$ occurs.'

Consider an example. We should all agree that I make the electrons in my

nerves move about in a certain way by wiggling my toes, and there is no mystery about this. But if a tragic accident severs my spinal cord and I can no longer wiggle my toes, we think the the reason for this is that certain of my nerves no longer allow electrons to move through them in the way needed for me to wiggle my toes. Indeed, if I can wiggle my toes I can make my the electrons in my nerves move in the needed way by wiggling my toes; but if after the accident the electrons suddenly start properly flowing through my nerves again, we cannot dispel the apparent mystery in this by pointing out that I can wiggle my toes and that I make my the electrons in my nerves move about by wiggling them. We are mystified by my miraculous toe-wiggling because we are mystified at how my nerves began working properly; if there is no account of the latter, there is no account of the former. And the best explanation for all this is that we think the motion of the electrons in my nerves are (part of) what makes my toes wiggle, and not *vice versa*.

Of course, if we deny that '*S* does *a* by *b*-ing' means 'the occurrence of *b* is part of what makes it the case that *a* occurs', we will want some other account of what this locution means. But such an account, perfectly consistent with the above observation, does not seem hard to find. At a first pass, we can say that, when *S* does *a* by *b*-ing, then *a* is one of the consequences of *S*'s *intention* to *b*. When I make my nerves work by wiggling my toes, for instance, my intention to wiggle my toes is what causes (and thus a good part of what makes) my nerves to work in a certain way, and the nerves working in that way is a good part of what makes it the case that my toes wiggle, and still say that I make my nerves work correctly by wiggling my toes.

In this case, though, it is clear that my control over the electrons in my nerves is nothing like direct control: I have control over the electrons in my nerves because I have (perhaps direct) control over something else — my intentions — which is (if my nerves are working properly) part of what makes it the case that those electrons move about as they do. Furthermore, *without* such an explanation, it would be mysterious *both* how I was ever able to make those electrons move and also how I was ever able to wiggle my toes.

So the objection really lends no plausibility to the thought that *S* has direct control over *X* — it suggests only that, if *X* has control over *R*, *S* also has control over *X*. But given Naturalism's thesis (N3), *S* cannot have (derivative) control over *X* by virtue of having direct control over R. (N3) tells us that the supervenience base of an action (a macrophysical event) must be what makes it the case that the event occurs, and not vice versa. This, combined with (B), tells us that *S* can have control over *R* only if she has control over the events in *r*'s super-

venience base — and given the rest of the argument, $x$ is the only reasonable candidate. But we have already seen that $S$ cannot have control over anything in $x$'s causal history; so $S$ cannot have control over $X$ whatsoever, and N$X$ is true.

# III Implications

The Supervenience Argument was presented above involving a particular act with a particular causal history, but its generality should be clear. We of course have no idea what the actual causal history of the first (allegedly) choosy act is like. If we subscribe to naturalism,[19] though, we will be committed to the supervenience of the first choosy act on some set of microphysical events. Furthermore, there will be some microphysical causal chain running from the supervenience base of whatever caused the action to the supervenience base of the action, and some of the events in the chain will be indeterministic while others won't be. If we let $IN$ express the occurrence of all the undetermined elements in the causal chain and $X$ express the occurrence of all the undetermined elements in the action's supervenience base we can use ($\beta\Box$) to generate essentially the same argument for any alleged first choosy action.

If this is right, what might we conclude? I can see six options:

(1) That the argument is sound and we don't have free will.

(2) That naturalism is false.

(3) That ($\beta\Box$) or ($*$) is valid but (B) is invalid.

(4) That (B), ($\beta\Box$), and ($*$) are all invalid.

(5) That the argument is sound, but choosy actions are not required for free will, so we are free (and (T1) is false).

(6) That we have direct control over microphysical actions.

To conclude either (1) or (2) (or their disjunction) would be, I take it, to acquiesce: free will is incompatible with naturalism. (3) is unmotivated: if ensuring the falsity of $p$ doesn't require ensuring the falsity of *what makes p true*, why should it require ensuring the falsity of *other* things that entail $p$? I can think of no principled reason to accept (3) other than that such an acceptance saves the Consequence Argument while undermining the Supervenience Argument. But if so, an appeal to (3) would be *ad hoc*.

[19] Since any stronger version of the naturalistic thesis entails the weak version, I purposely leave the claim ambiguous.

To accept either of (4) and (5) would be to give up on the Consequence Argument. Clearly, if ($\beta\Box$) is rejected, the argument is flatly invalid. Likewise, if ($*$) is rejected, although the Consequence Argument retains its validity, the support for the second premiss, $\mathrm{N}(P \,\&\, L)$, is lost. And if (T1) is rejected, then both the Consequence and the Supervenience Arguments become irrelevant: even if they successfully show that nobody ever acts choosily, they leave wide open the possibility that people often act freely.[20]

This leaves only (6), the claim that we have direct control over microphysical events. I find (6) incredibly implausible. At the very least, it runs against the spirit, if not the letter, of naturalism, in which case a defence of the compatibility of the two based on it seems rather pointless.

The situation for (6) is actually rather worse than this. But before explaining why, it will help to say a bit about the dialectical purpose the Supervenience Argument is supposed to serve. It occupies a position in rhetorical space similar to that of (what van Inwagen calls the third strand of) the *Mind* argument. The *Mind* argument was supposed to show that, if ($\beta$) is valid, then free will is also incompatible with indeterminism (van Inwagen 1983: 142–150). According to the argument, if indeterminism is true, then if an action $r$ has a particular set of indeterministic causes $db$, nobody could have done anything to ensure that $db$ caused $r$. Thus, $\mathrm{N}(DB \rightarrow R)$ is true. And, if $r$ is the first choosy act, then $\mathrm{N}DB$ is true.[21] A single application of ($\beta$), however, yields the conclusion that $\mathrm{N}R$ is true as well. Thus $r$ is not a choosy act after all, and so choosy acts and therefore free will are non-existent. Indeterminism precludes free will. Thus, libertarians have as much reason to reject ($\beta$) (or (T1)) as compatibilists do.

The *Mind* argument was supposed to place a certain sort of pressure on libertarians: give up on the Consequence Argument or give up on free will. Finch and Warfield (1998) argued that the situation changed once independent counterexamples to ($\beta$) came along. But Dana Nelkin (2001) has shown us how to revise the argument in light of ($\beta$)'s invalidity. Suppose $r$ is the first choosy action and is indeterministically caused by $db$. Now consider the propositions $DB$ and $DB \rightarrow R$. Since $DB$ is in the broad past of the first choosy action, there is nothing anyone could do that even might have ensured its falsity, and so by

[20] (4) and (5) don't exhaust the ways we could reject both the Supervenience and Consequence Arguments together. We could, for instance, follow the 'Humean Compatibilists' (Beebee and Mele 2002) and deny $\mathrm{N}L$. But the general point should be clear: many ways of resisting the Supervenience Argument are just as threatening to the Consequence Argument.

[21] I take the liberty of making the Mind argument about the first choosy act, but I do it only because it strengthens the plausibility of the first premiss.

$(*)$, if N$DB$ and N$(DB \rightarrow R)$ are true, N$(DB \,\&\, (DB \rightarrow R))$ will be also.[22] The argument then runs:

**Nelkin's Revised *Mind* Argument**

| | |
|---|---|
| (1) N$(DB \,\&\, (DB \rightarrow R))$ | Premiss |
| (2) $\Box((DB \,\&\, (DB \rightarrow R)) \rightarrow R)$ | Logical Truth |
| (3) N$R$ | ($\beta\Box$: 1, 2) |

The *Mind* Argument thus survives ($\beta$)'s invalidity.

If it could also perform its proper dialectical function, the Supervenience Argument might not be worth the bother. But regardless of its *validity*, it is not at all clear why we should accept the premiss N$(DB \rightarrow R)$ of the original argument, and thus not clear why we should accept the first premiss of the revised Argument. Notice that $DB \rightarrow R$ is truth-functionally equivalent to $\sim DB \vee R$. Since the first disjunct is false, it would seem that all $S$ would need to do to ensure the entire disjunction's falsity would be to refrain from $r$-ing. But if $R$ expresses the occurrence of an action, why couldn't the agent have direct control over $R$? So it seems then that we have no reason to accept N$(DB \rightarrow R)$ *other than* a prior commitment to N$R$.[23]

So the *Mind* Argument does not seem well-equipped to provide the kind of pressure it is supposed to provide; and where the *Mind* argument fails, the Supervenience Argument can step in to place renewed pressure on libertarians. This pressure is limited, though, to those libertarians who are also naturalists; non-naturalistic libertarians are free to simply reject the assumptions that generate certain premisses of the Supervenience Argument. But if the pressure is limited in this way, is the Supervenience Argument even worth bothering with?

Yes. First, it is not insignificant that naturalistic libertarians cannot use the Consequence Argument to support their position. Some of the most popular libertarian positions on the table (e.g., that of Robert Kane 1996, 1999) are explicitly naturalistic, and in their wake, other attempts to secure libertarian free will in a naturalistic framework (e.g., that of Laura Waddell Ekstrom 2000) have been

[22] This is not quite Nelkin's argument, which relied on N$(DB \rightarrow R)$ also being in $r$s broad past, which looks contentious. Thanks to an anonymous referee for helping me see that an appeal to $(*)$ could circumvent any need for this contentious claim. (Cf. also Carlson 2002)

[23] Perhaps this is too strong; van Inwagen (1983: 142–145), for instance, seems to provide some argument for the premiss. At any rate, though, there are ways to resist the *Mind* argument — such as arguing that there's no reason indeterminism alone should think we have no choice about certain propositions — which do not apply to the Supervenience Argument.

developed. But many of these libertarians explicitly rely on the Consequence Argument to defend their incompatibilism (cf. Ekstrom 2000: 26–42),[24] and even those who don't are often committed to the claim that, if agents have free will of the sort their theories calls for, then the Consequence Argument shows that determinism precludes a necessary condition for free will (cf. Kane 1996: 75–77). If the success of the Consequence Argument as an argument for incompatibilism entails the incompatibility of free will and naturalism, then many well-received theories of libertarian free will fall to internal inconsistency.

Second, there is a *reason* these naturalistic views are so popular. In the philosophical arena, positions are evaluated on a number of merits, one of which is overall plausibility. Naturalism is considered by many to be an extremely plausible view, and while incompatibility with it may not count decisively against a position, it is a cost to be avoided. Naturalistic libertarianism, by making use of the indeterminism ready-to-hand in the natural order, thus comes cheap. Non-naturalistic libertarianism, by contrast, is costly. For some, at least, that cost will seem high enough to warrant a rejection of either the Consequence Argument or free will rather than payment.

This point can be put another way. Van Inwagen (1992: 58) has suggested that the debate between compatibilists and incompatibilists be thought of as taking place in front of an audience as of yet undecided about whether free will is compatible with determinism. The goal of each debater is not the overly ambitious one of convincing her opponent, but instead the more conservative one of converting the agnostic audience to her position. Presumably, the debate between libertarians and their opponents (compatibilists and those sceptical about free will) should be viewed the same way: the libertarian is trying to convince not her opponents, but the agnostic audience both that we have free will and that it is incompatible with determinism.

As already noted, naturalism is a widespread view. Many in the agnostic audience are liable to resist the thought that free will has to be sought outside that part of the world open to scientific investigation. Many philosophers become sceptical when their colleagues begin to 'look for [free will] in mysterious sources outside of the natural order or to postulate unusual forms of agency or causation' (Kane 1996: 115). Naturalism, for better or for worse, is a widely held

[24] Ekstrom prefers a version of the Consequence Argument that uses the operator '$\mathrm{B}_{S,t}$', where '$\mathrm{B}_{S,t}p$' means '$p$ and $S$ is not able at $t$ to prevent its being the case that $p$' (2000: 28–29). Ekstrom endorses the relevant analogues to ($\alpha$), ($\beta$), and ($\beta\Box$); since premiss (2) of the tricky argument seems no less plausible when 'N' is replaced with '$\mathrm{B}_{S,t}$', the Supervenience Argument threatens her position as well.

view, and people do not want to sacrifice it on the free-will altar if an alternative can be found.

This means that a defence of libertarianism *in general* is much more difficult. Before, even a non-naturalistic libertarian might have been able to convince many in the agnostic audience to at least be *libertarians* even if they would not go all the way with her into non-naturalism. Now, however, convincing the audience that libertarianism is warranted requires convincing them that the high non-naturalistic costs of libertarianism are both necessary for free will and worth paying.

In light of the argument's dialectical function, we can see why an appeal to (6) is not an incredibly fruitful line for the naturalistic libertarian to pursue. Just as an agnostic audience who wants to locate free will in the natural order will not be excited about rejecting naturalism to secure free will, they will not be excited about the thought that we have *direct* control over the microphysical, either.

We can, as (van Inwagen 1983: 95–96) does, think of the dialectical situation in terms of how plausible we find each reaction to the argument. Of (1)–(6), which is most plausible, and which is least? For my part, option (6) sits towards the bottom of the plausibility scale. And I suspect that the vast majority of the agnostic audience — or, at least, the vast majority of that portion of the audience that already leans toward naturalism — will agree with me on the relative plausibility of (6). So they will be almost certain to find some other response — a rejection of the Consequence Argument or a rejection of naturalistic libertarianism — a more attractive prospect.

Nonetheless, some will prefer to hang on to naturalistic libertarianism with (6). And I have no argument that will force them not to. But even so, let it not be thought that the Supervenience Argument is thereby uninteresting; at least *some* who once thought naturalistic libertarianism viable will also balk at the thought that we have direct control over microphysical events. Furthermore, even naturalistic libertarians who embrace (6) may well have thought beforehand that they could get by without it. Naturalistic libertarianism did not come obviously pre-packaged with a commitment to direct control over microphysical events, and its proponents may be surprised to discover that they need it to keep the view viable. Even those willing to pay the price ought to find it significant that the price indeed needs paying.[25]

[25]Thanks to Zac Ernst, Peter Hanowell, Matt James, Jeremy Kirby, Kirk Ludwig, Christopher Pynes, and especially Tom Crisp, Al Mele, and Eddy Nahmias, for helpful discussion and comments. Thanks also to two anonymous referees for some excellent comments, criticisms, and

# References

Beebee, Helen and Alfred R. Mele (2002). "Humean Compatibilism." *Mind* 111: 201–223.

Bishop, John (2003). "Prospects for a Naturalist Libertarianism: O'Connor's *PersonsandCauses*." *Philosophy and Phenomenological Research* 66: 228–243.

Bricker, Phillip (2006). "Absolute Actuality and the Plurality of Worlds." *Philosophical Perspectives* 20: 43–76.

Campbell, Joseph Keim (2007). "Free Will and the Necessity of the Past." *Analysis* 67(2): 105–111.

Carlson, Erik (2000). "Incompatibilism and the Transfer of Power Necessity." *Noûs* 34: 227–290.

— (2002). "In Defence of the *Mind* Argument." *Philosophia* 29: 393–400.

Clarke, Randolph (2003). *Libertarian Accounts of Free Will*. New York: Oxford University Press.

Cover, J. A. and John (O'Leary) Hawthorne (1996). "Free Agency and Materialism." In Jeff Jordan and Daniel Howard-Snyder (eds.), *Faith, Freedom, and Rationality*, 47–71. Lanham, Md.: Rowman & Littlefield.

Crisp, Thomas M. and Ted A. Warfield (2000). "The Irrelevance of Indeterministic Counterexamples to Principle Beta." *Philosophy and Phenomenological Research* 61: 173–184.

Ekstrom, Laura Waddell (2000). *Free Will: A Philosophical Study*. Boulder, Co.: Westview Press.

Finch, Alicia and Ted A. Warfield (1998). "The *Mind* Argument and Libertarianism." *Mind* 107: 515–528.

Ghirardi, Gian-Carlo, Alberto Rimini and Tuillio Weber (1986). "Unified Dynamics for Microsopic and Macrosopic Systems." *Physical Review* 34: 470–491.

Ginet, Carl (1990). *On Action*. New York: Cambridge University Press.

suggestions that greatly improved the final product.

Huber, Daniel, Leopoldo Petreanu, Nima Ghitani, Sachin Ranade, Tomás Hromádka, Zach Mainen and Karel Svoboda (2007). "Sparse Optical Microstimulation in Barrel Cortex Drives Learned Behaviour in Freely Moving Mice." *Nature* 451(3): 61–63.

Huemer, Michael (2000). "Van Inwagen's Consequence Argument." *The Philosophical Review* 109(4): 525–544.

Kane, Robert (1996). *The Significance of Free Will*. New York: Oxford University Press.

— (1999). "Responsibility, Luck and Chance." *The Journal of Philosophy* 96: 217–240.

Kim, Jaegwon (1984*a*). "Concepts of Supervenience." *Philosophy and Phenomenological Research* 45: 153–176. Reprinted in Kim (1993): 53–78.

— (1984*b*). "Epiphenomenal and Supervenient Causation." *Midwest Studies in Philosophy* 9: 257–270. Reprinted in Kim (1993): 92–108.

— (1987). "'Strong' and 'Global' Supervenience Revisited." *Philosophy and Phenomenological Research* 48: 315–326. Reprinted in Kim (1993): 79–91.

— (1993). *Supervenience and Mind*. Cambridge: Cambridge University Press.

Loewer, Barry (1996). "Quantum Mechanics and Free Will." *Philosophical Topics* 24: 91–112.

McKay, Thomas and David O. Johnson (1996). "A Reconsideration of an Argument Against Compatibilism." *Philosophical Topics* 24: 113–122.

Merricks, Trenton (2001). *Objects and Persons*. Oxford: Clarendon Press.

Nelkin, Dana K. (2001). "The Consequence Argument and the *Mind* Argument." *Analysis* 61: 107–115.

O'Connor, Timothy (1993). "On the Transfer of Power Necessity." *Noûs* 27: 204–218.

— (2000). *Persons and Causes*. New York: Oxford University Press.

Schaffer, Jonathan (2000). "Causation by Disconnection." *Philosophy of Science* 67(2): 285–300.

Stalnaker, Robert (1996). "Varieties of Supervenience." *Philosophical Perspectives* 10: 221–241.

Unger, Peter (2002). "Free Will and Scientiphicalism." *Philosophy and Phenomenological Research* 65: 1–25.

van Inwagen, Peter (1983). *An Essay on Free Will*. Oxford: Oxford University Press.

— (1992). "Reply to Christopher Hill." *Analysis* 52: 56–61.

— (2000). "Free Will Remains a Mystery." *Philosophical Perspectives* 14: 1–19.

Warfield, Ted A. (2000). "Causal Determinism and Human Freedom are Incompatible: A New Argument forIncompatibilism." *Philosophical Perspectives* 14: 167–180.

Widerker, David (1987). "On an Argument for Incompatibilism." *Analysis* 47: 37–41.