The Incompatibility of Free Will and Naturalism*

Jason Turner

September 26, 2007

The *Consequence Argument* is a staple in the defense 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, that libertarian 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, since 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.

1 The Consequence Argument

The Consequence Argument purports to show that free will is incompatible with determinism, where the latter thesis is understood as the claim that the laws of nature, conjoined with any proposition accurately describing the entire state of the world at some given time, entail any other true proposition.[1] 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, p. 56).

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 "Np" means "p, and no one has, or ever had, any choice

*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.

[1] Cf. van Inwagen 1983, pp. 58–65. Note that p's entailing q is here understood as the necessity of the conditional $p \rightarrow q$.

about whether p (van Inwagen 1983, p. 93).[2] (I will follow Alicia Finch and Ted A. Warfield (1998, p. 516) in understanding "someone has a choice about whether p" as "someone could have acted so as to ensure the falsity of p.") Originally, the Consequence Argument appealed to two inference rules:

(α) From $\Box p$, deduce Np, and

(β) From Np and N$(p \rightarrow q)$, deduce Nq

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

(Agg) From Np and Nq, 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.[3]

In response, proponents of the Consequence Argument have rejected (α) and (β) in favor of another inference rule which does not entail (Agg):

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

[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 ϕ a sentence expressing p, Np 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; and McKay and Johnson 1996, note 4).

[3]In fact, a similar counterexample can be used to directly show the invalidity of (β). 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 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, pp. 283–284; Crisp and Warfield 2000, pp. 178–179. See Widerker 1987 for an independent counterexample to (β), and O'Connor 1993, p. 209 and McKay and Johnson 1996, pp. 117–118 for criticisms of Widerker's example.)

(Finch and Warfield 1998, pp. 521–522; Widerker 1987, p. 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)	$\mathrm{N}(P \,\&\, L)$	Premise
(2)	$\Box((P \,\&\, L) \rightarrow F)$	Premise (Determinism)
(3)	$\mathrm{N}F$	$\beta\Box$: 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. It seems intuitive that, as a result, we have no choice about the conjunction of these two propositions.

Since, as we have already seen, the N-operator is not agglomerative, we cannot argue for the truth of $\mathrm{N}(P \,\&\, L)$ by appealing to the truth of both $\mathrm{N}P$ and $\mathrm{N}L$. 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 argue:

> [T]he core intuition [described above] motivates the acceptance of [the first] premise. This core intuition is, we maintain, the intuition that the past is fixed and beyond the power of human agents to affect in any way. P describes the state of the world at some time in the distant past (before any human agents existed). L is a conjunction of the laws of nature which, we presume, in addition to being inalterable by human agents, do not change over time. Thus the conjunction $(P \,\&\, L)$ offers a description of what might be called the "broad past" — the complete state of the world at a time in the distant past including the laws of nature. We maintain, in asserting our premise, that the broad past is fixed [in a way that justifies $\mathrm{N}(P \,\&\, L)$] (1998, p. 523).

Thus we need not appeal to agglomeration to justify the contention that $\mathrm{N}(P \,\&\, L)$ is true given our intuitions that $\mathrm{N}P$ and $\mathrm{N}L$ are true, because those intuitions *directly* support the claim that $\mathrm{N}(P \,\&\, L)$ is true.[4]

2 Naturalism and the Supervenience Argument

There is a view about the nature of reality, which I will tag with the over-worked name of *naturalism*, that in rough form holds that everything eventually boils down to fundamental physics. This is not necessarily a reductionistic view (although global reductionism is one variant of naturalism), but rather a supervenience thesis which holds that all events and causal relations supervene on microphysical events and causal relations. As I understand it here, naturalism comprises two main claims:

(N1) All events supervene on microphysical events, and

(N2) All causal relations supervene on microphysical causal relations.

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.[5]

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

[4]Retreat to ($\beta\Box$) is not the only way to revise the Consequence Argument in light of (β)'s invalidity; we might instead offer a different interpretation of "N." For example, McKay and Johnson (1996, pp. 118–120) consider an interpretation on which "Np" means "p and no one could have acted in a way that *might* ensure the falsity of p" (see also Finch and Warfield 1998, pp. 524–527). 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 do not argue for it here, I think it can be shown that one should find any of these revisions of the Consequence Argument compelling only if one finds the ($\beta\Box$) version compelling; for this reason, I do not consider them in the text. Cf. Carlson (2000).

[5]If there are no such things as omissions — in other words, if events aren't closed under (the event counterpart of) negation — then this claim will be strictly stronger than standard formulations of supervenience. I assume, however, that events are closed under (the event counterpart) of negation (see the discussion of thesis (T3) below), so this claim follows from standard supervenience. See Kim 1984*a*, p. 64, Kim 1987, pp. 81–82, and Bricker 2006, pp. 267–270.

with divergent events have either divergent microphysical events or divergent laws of nature.[6]

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*, p. 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 this diluted sort of causal supervenience with merely nomic supervenience of events on the microphysical *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, claims to be consistent with (N2), since they both posit forms of substance causation not meant to supervene on microphysical event-causation.[7] As I use "naturalism" here, I mean to rule agent-causal theories, including O'Connor's 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, p. 115), no matter how friendly to current science those theories may otherwise try to be.[8]

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

[6]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, pp. 228–230).

[7]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. In order to accommodate this potential position, (N2) should probably be augmented with the claim "and all causation is event-causation." Thanks here to Tom Crisp.

[8]According to J. A. Cover and John O'Leary-Hawthorne (1996), agent-causal theories that subscribe to (something like) (N1) but not (N2) are coherent but implausible. I remain neutral about that claim here.

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 favor in the eyes of some. Trenton Merricks, for instance, gives what I take to be a version of it in *Objects and Persons* (2001, pp. 155–161). Others have given related arguments against the compatibility libertarian free will and a broadly naturalistic worldview (e.g. Bishop 2003; Loewer 1996; Unger 2002). However, no one has taken the trouble to clothe the Supervenience Argument in the same formal garb the Consequence Argument wears. 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. The remainder of this paper provides this formal clothing and considers the nature of the ties between these two arguments.

2.1 Choosy Actions

Call an event *a choosy* if and only if $A \;\&\; \sim \mathrm{N}A$ is true, where A is a proposition expressing the occurrence of a.[9] The Supervenience Argument proceeds as follows. 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 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.

[9] 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.

What should we say about these theses?

Let's begin with (T2). How could it be false? Here's one way: there could be an infinite number of choosy actions, each one preceded by an earlier one. Here's another: there could be a tie for the first choosy action. I can think of no other way (T2) could be false.

I want to postpone consideration of the backwards-infinite cases until later. Let's think just about ties here.[10] Here's one way there could be a tie: a "fine-grained" theory of actions (and events more generally) might be true, in which case, if Sam choosily raises his arm at t and nobody ever acts choosily before t, there might in fact be many distinct choosy actions that Sam performed at t (including, for instance, raising his right arm, raising his arm rapidly, raising his arm in the presence of a veiled assassin, etc.).

If a fine-grained theory of action is true, then (T2) is not. We can modify the Supervenience Argument for use with the fine-grained theory, though. According to a coarse-grained theorist, when Sam raises his arm at t, he performs one action that we could describe in many ways. According to the fine-grained theorist, each of these descriptions (if they are not logically equivalent) describes a distinct action. The fine-grained theorist will concede, however, that these actions are all related to each other in a way that they are not related to any other actions.[11] So every action has an equivalence class given by this relation, and this equivalence class is a sort of ersatz coarse-grained action. Call such an equivalence class of actions choosy if and only if it contains at least one choosy action, and say that one such class is "before" another if the former includes some action that is occurring before any actions in the latter are occurring. Then we can replace (T2) with

> (T2′) If there are some choosy equivalence classes of actions, there is a first one,

and use the argument I give below to show that *every* action in the first choosy equivalence class is not choosy. Thus the class is not choosy, so there are no choosy equivalence classes of actions and therefore no choosy actions. (Having suggested the modification, I set it aside, speaking with the coarse-grained theorist throughout the paper unless otherwise noted.)

[10] I set aside concerns about general relativity; assume we've fixed a frame of reference F and are talking about the "first" choosy action with respect to F.

[11] For instance, they are all on the same "action tree" as a. Very roughly, an action tree contains all actions performed by a single agent that are related to each other via the "by"-relation in the right sort of way. For instance, if Joe pleases Mary at 10:45 by purchasing her some flowers at 9:00, then the actions described by "Joe pleases Mary at 10:45," "Joe purchases some flowers at 9:00," and "Joe purchases some flowers at 9:00 while thinking about Mary" are all on the same action-tree (see Ginet 1990, pp. 19, 46–52).

There could also be "simple" ties. Perhaps Sam and Max both perform choosy actions at t and nobody performs any choosy actions before t. This would also falsify (T2). Since space-like spearated events cannot be causally relevant to each other,[12] we can arbitrarily pick one of these to be the "first." More generally, we can order the set of choosy actions in a way that, if actions x and y both occur at t, one will count as "before" the other thanks to their spatial relations (to an arbitrary point and from an arbitrary direction, say). Assuming that two actions (or equivalence classes of actions) cannot fully spatiotemporally coincide, there will be no ties in *this* competition.

I thus tentatively endorse (T2), with these modifications at the ready. What about (T1)? Call a proposition choosy if and only if $p \,\&\, \sim \mathrm{N}p$ is true. If the Consequence Argument is to be of any force, choosy propositions must be a necessary condition on free will. Choosy actions will also be necessary if the following is true:

> (T3) There are choosy propositions if and only if there are choosy actions.

The right-to-left direction of this claim is trivial: there is a choosy action if and only if the proposition expressing its occurrence is choosy. But could there be choosy propositions but no choosy actions?

Suppose p is choosy; thus there is something someone could have done to ensure the falsity of p. Since p is true, they did not do that thing. But perhaps they did not perform any *other* action in its stead, either. Perhaps, in fact, for every choosy proposition p, whoever could have acted so as to ensure p's falsity instead merely failed to act. Then there could be free will without choosy actions.

If omissions are a species of action, then the above story is incoherent; failing to act to ensure p's falsity is itself an act. If omissions are not actions but are another species of event, then the above story is coherent, but we can simply revise (T1)-(T3), and the Supervenience Argument, in terms of events rather than actions.

Suppose omissions are neither actions nor events. Or, more precisely, suppose there is no species of entity rightly called "omissions." Even so, if we can make sense of talk of possible actions, we can find a truth in the vicinity of (T1). For, if there is a choosy proposition p, there is at least a *possible* action a that someone could have performed and which, if they had performed it, would have ensured the falsity of p. So we could rewrite (T1)–(T3) in terms of possible actions rather than (actual) actions to avoid the counterexample.

[12] Assuming all causation is local, at any rate; I cannot see how non-locality could possibly help free will in a naturalistic universe

Finally, even if we eschew possible actions, the needed theses can be recovered if we think propositions exist necessarily. If I could have performed an action, then if I had done so, there would have been a singular proposition q that expressed the occurrence of that action. If q would have existed necessarily, then q actually exists (assuming S5 axioms hold for metaphysical necessity). In this case, its negation exists, and presumably is made true when I fail to act. So we can rewrite theses (T1)–(T3) in terms of action-expressing propositions and revise the arguments below accordingly.

We can carry these considerations further, but it looks clear that denying the existence of omissions will not bar us from finding *some* type of ersatz action with which we can define a suitable notion of choosyness and recast the Supervenience Argument. I will suppose for the balance of the paper that (T1)–(T3) are true, keeping in mind that, if they are not, the arguments I give may be modified along one of the lines suggested here.

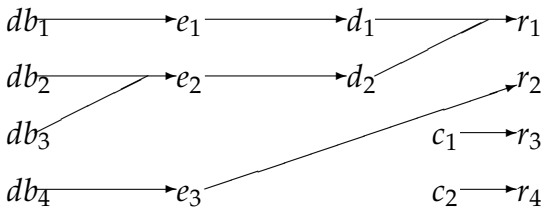
2.2 Causal Relations

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. 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. For the sake of exposition, I will suppose they each have only four events in their supervenience base ($db_1, \ldots, db_4$ and $r_1, \ldots, r_4$, respectively); a more realistic count would be unworkable. 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.

Suppose for illustration that the causal chains work like this: db_1 causes an event, e_1, which in turn causes another event, d_1. 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

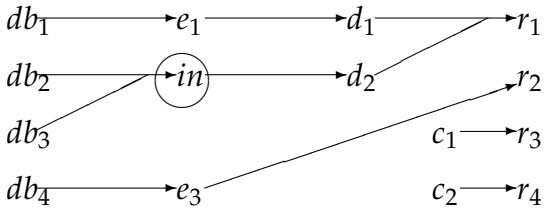
of r_3 and r_4, respectively. The causal chain between db^+'s supervenience base and r's supervenience base will then look like this:



Now, if this is a deterministic universe, the occurrence of $db_1, \ldots, db_4, c_1$, and c_2 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.



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, …, R_4 propositions expressing the occurrence of each of r_1, …, r_4, respectively, then the following proposition is true:

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

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

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

2.3 The Supervenience Argument: A First Pass

We are now in a position to offer a formal version of the Supervenience Argument. The first premise 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. It should be clear how the argument is supposed to go:

The Supervenience Argument:

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

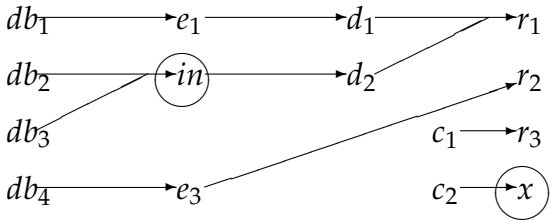
The second premise seems to follow from the "broad past" principle Finch and Warfield used to defend the Consequence Argument. In that instance, the intuitions supporting the claim that $\mathrm{N}(P \,\&\, L)$ is true were that both P and L were true long before there were any humans around and that the past is fixed. Apparently, the idea is that, since P & L was true before anyone could have done anything to falsify it, and since we cannot *now* do anything to falsify what has gone on before, nothing we can now do could falsify P & L.

Similar reasoning lends support to the claim that $\mathrm{N}(DB \,\&\, C \,\&\, IN \,\&\, L)$ is true. The proposition $DB \,\&\, C \,\&\, IN \,\&\, L$ is made true *before r occurs*, and r is the first choosy act. One might object that, even though r is the first choosy *action*, there are choosy propositions that occur before r. However, the sort of reasoning used in defense of (T3) above also applies here: if some proposition p in the past of r is choosy, then there is some action someone could have performed, but didn't, that would have ensured the falsity of p. Assuming (as I am, with fallback positions in hand) that omissions are actions, this omission would have been a choosy action that occurred before r; thus r would not have been the first choosy action after all.

2.4 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," as it were. But there may instead be a causal chain from the *db*'s 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 below.)



It appears that nobody can, or ever could have, ensured the falsity of X, the proposition that expresses the occurrence of x. How could anyone exercise such control over the truly objective chance happenings of particle physics? What could I do, for instance, to ensure that an electron will have a certain property at a certain time, if it is *objectively undetermined* whether or not it will gain said property at said time? (Cf. Loewer 1996; van Inwagen 1983, pp. 142–143.)

As far as I can see, there is nothing I (or anyone) could do that would determine the outcome of an undetermined microphysical event. 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)	$\Box((DB \,\&\, C \,\&\, IN \,\&\, X) \rightarrow (L \,\&\, R_1 \,\&\, R_2 \,\&\, R_3 \,\&\, X))$	Premise (N2)
(2)	$\mathrm{N}(DB \,\&\, C \,\&\, IN \,\&\, L \,\&\, X)$	Premise
(3)	$\mathrm{N}(L \,\&\, R_1 \,\&\, R_2 \,\&\, R_3 \,\&\, X)$	$\beta\Box$: 1, 2
(4)	$\Box((L \,\&\, R_1 \,\&\, R_2 \,\&\, R_3 \,\&\, X) \rightarrow R)$	Premise (N1)
(5)	$\mathrm{N}R$	$\beta\Box$: 3, 4

The first premise 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 premise 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. The problem is that I cannot use agglomeration to support the second premise, and x does not lie in the "broad past" of r.

Nonetheless, I claim that $\mathrm{N}(DB \,\&\, C \,\&\, IN \,\&\, L \,\&\, X)$ is true. According to Finch and Warfield,

> ...it is important to be clear that the McKay and Johnson argument [against agglomeration] shows only that the inference from $\mathrm{N}p$ and $\mathrm{N}q$ to $\mathrm{N}(p \,\&\, q)$ is invalid. This does not, by itself, provide any reason at all for thinking that [in the case of $\mathrm{N}P$ and $\mathrm{N}L$] $\mathrm{N}P$ and $\mathrm{N}L$ are true, while $\mathrm{N}(P \,\&\, L)$ 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 $\mathrm{N}(P \,\&\, L)$. 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, pp. 523-524, emphasis added)

Similar remarks apply here. There does not seem to be anything one could do that even *might* ensure the non-occurrence of x, the truly undetermined event, or *might* ensure the falsity of true propositions about the past, and by virtue of this

would ensured the falsity of DB & C & IN & L & X. Premise (2) is vindicated and the argument follows.

Some may object: since *in* and *x* are both undetermined, perhaps some action *a* that *S* did in fact perform, 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 is not enough to show that (2) is false. 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 there is something they could have done which *would* have ensured the falsity of p & q. Suppose Herbert tosses a six-sided die; his toss might ensure the falsity of *the die does not land one*, and it might ensure the falsity of *the die does not land six*, but it's just wrong to say that his toss *would* ensure the falsity of *the die does not land one* & *the die does not land six*. 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 not falsified.

3 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,[13] 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 indeterministic elements in the causal chain and X express the occurrence of all the indeterministic 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, anyone who finds the Consequence Argument a persuasive argument for incompatibilism should also hold that the existence of choosy actions — and therefore the existence of free will — requires the falsity of naturalism. In the following section, I defend the Supervenience Argument against certain objections. In this section, I consider the ramifications if that defense is successful.

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

The Supervenience Argument occupies a position in rhetorical space similar to that of the *Mind* argument. The *Mind* argument was supposed to show that, if (β) is valid, then free will is also incompatible with *indeterminism* (van Inwagen 1983, pp. 142–150; note that van Inwagen calls this the "third strand" of the *Mind* argument). 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.[14] A single application of (β), 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 (β) as compatibilists do.

Before the advent of independent counterexamples to (β), the *Mind* argument placed a lot of pressure on libertarians. They could not use their favorite argument against compatibilists without opening themselves up to charges of self-contradiction. Once it was seen that (β) was invalid, though, the *Mind* argument lost most of its force. Libertarians could now happily reject (β) — and the *Mind* argument with it — while weilding a Consequence Argument patched up with ($\beta\Box$) (This is the main point of Finch and Warfield 1998).

Of course, if the *Mind* argument could also be reworked to use ($\beta\Box$), libertarians would once again be in a bind. But it cannot. Dana Nelkin (2001) has given the best attempt at such a reconstruction. She argues that, if r is the first choosy action and is indeterministically caused by db, then both DB and $DB \rightarrow R$ lie in the "broad past" of r. She then reformulates the argument as follows:

Nelkin's Revised Mind Argument:

(1)	$\mathrm{N}(DB \ \& \ (DB \rightarrow R))$	Premise
(2)	$\Box((DB \ \& \ (DB \rightarrow R)) \rightarrow R)$	Logical Truth
(3)	$\mathrm{N}R$	$\beta\Box$: 1, 2

Notice that $(DB \ \& \ (DB \rightarrow R))$ is truth-functionally equivalent to $(DB \ \& \ R)$, so at first blush it appears too strong a premise for libertarians to accept. As O'Connor (2000, note 12) observes, it would be "an unusually inept compatibilist" that would accept a Consequence Argument requiring $\mathrm{N}[(P \ \& \ L) \ \& \ ((P \ \& \ L) \rightarrow F)]$ as a premise, since they would be in essence granting the truth of $\mathrm{N}((P \ \& \ L) \ \&$

[14] Making the *Mind* arguments (both this one and Nelkin's revised version below) about the first choosy act is a liberty I have taken, but it is a liberty that strengthens the arguments.

F). Likewise, it would be an unusually inept incompatibilist who would accept (1), at least without argument, for that comes uncomfortably close to simply granting that NR is true.

Of course, Nelkin has an argument: both DB and $DB \rightarrow R$ are supposed to lie in the "broad past" of r (2001, p. 113). But this claim appears false. In particular, in an indeterministic world where DB is true, $DB \rightarrow R$ — a notational variant of $\sim DB \vee R$ — appears to be made true no earlier than R is. It is thus difficult to see how it could then lie in r's "broad past." Unless a better argument is forthcoming for (1), the ($\beta\Box$)-revised *Mind* argument appears to have lost.[15]

In the absence of a working ($\beta\Box$)-*Mind* argument, the Supervenience Argument can put a similar sort of pressure on libertarians: give up ($\beta\Box$) or give up on free will in order to remain consistent. This pressure is limited, however, to libertarians who are also naturalists; non-naturalistic libertarians are free to simply reject the assumptions that generate certain premises of the Supervenience Argument. If the pressure is limited in this way, though, then 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. The most popular libertarian position on the table — that of Robert Kane (1996; 1999) — is explicitly naturalistic, and in its wake, other attempts to secure libertarian free will in a naturalistic framework — such as that of Laura Waddell Ekstrom (2000) — have been proposed. Kane, however, is committed to the claim that if agents have free will of the sort his theory calls for, then the Consequence Argument shows

[15]Erik Carlson (2002) offers his own reformulation of the *Mind* argument, which does not use ($\beta\Box$). It argues from the truth of $\mathrm{N}_{S,t}DB$ and $\mathrm{N}_{S,t}(DB \rightarrow R)$, where "$\mathrm{N}_{S,t}p$" means "$p$ and agent S has at time t no choice about whether p" (where "having a choice" is understood in the way suggested by Finch and Warfield (1998, p. 516)), to the truth of $\mathrm{N}_{S,t}R$, via the principle

> (β^{**}) From (i) $\mathrm{N}_{S,t}p$, (ii) $\mathrm{N}_{S,t}(p \rightarrow q)$, and (iii) the fact that, for any way in which S is able at t to act so that q would be false, either: if S were to act in that way, p would be false, or if S were to act in that way, p would be true, deduce $\mathrm{N}_{S,t}q$ (cf. Carlson 2002, p. 397).

For the argument, t is "the last time at which it is possible for S to exercise her (alleged) ability to choose whether or not [r] shall occur" (Carlson 2002, note 11).

It is reasonable to think that some of r's causes are still occurring after others have run their course. Suppose that d and b are two of r's causes, where d occurs at t and b occurs before t. (It is not implausible to think that some of r's causes will occur at the last moment at which S could have exercised her (alleged) ability to choose whether or not r would occur.) Suppose further that b causes r via a causal chain which includes e, which occurs at t. Suppose finally that both b and e are undetermined. Then there may be close worlds in which r does not occur because b doesn't occur (and hence in which DB is false) and equally close worlds in which r does not occur, b does occur, and e does not occur (in which DB is true). Condition (iii) of (β^{**}) would thus not be met for the relevant propositions, so the argument fails.

that determinism precludes a necessary condition for free will (1996, pp. 75–77), and Ekstrom relies primarily on the Consequence Argument to defend her incompatibilism (2000, pp. 26–42).[16] If the success of the Consequence Argument as an argument for incompatibilism entails the incompatibility of free will and naturalism, then two well-received theories of libertarian free will fall to internal inconsistency.

Second, there is a *reason* views like Kane's and Ekstrom's 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 ($\beta\Box$) or free will rather than payment.

This point can be put another way. Van Inwagen (1992, p. 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 skeptics) should be viewed the same way: the libertarian is trying to convince 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 skeptical 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, p. 115). Naturalism, for better or for worse, is a widely held view, and people do not want to sacrifice it on the free-will altar if an alternative can be found.

This means that a defense of libertarianism *in general* is much more difficult. Before, 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

[16] 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, pp. 28–29). Ekstrom endorses the relevant analogues to (α), (β), and $(\beta\Box)$; since premise (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.

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. This is likely to be a tough sell.

4 Objections and Replies

Naturalistic libertarians motivated by the Consequence Argument are in something of a bind, for they must find a way to reject the Supervenience Argument which does not in turn license a rejection of the Consequence Argument. In this section I respond to potential objections that try to circumvent the Supervenience Argument while leaving the Consequence Argument unscathed.

Objection 1: Suppose that the universe is such that, for every choosy act, there is an earlier choosy act. Then the set of all choosy actions would not have a first element, but it clearly would not be empty.

There are two ways a non-empty, linearly-ordered class of choosy actions could fail to have a first element. First, it could be that, for every time, there is an earlier time at which someone performed a choosy act. Call worlds like this *eternally choosy*. Second, there could be some last time t_0 before which nobody had ever performed a choosy act such that, for every time t later than t_0, there is a time t_- between t_0 and t at which there is a (distinct) choosy act. Call worlds like this *densely choosy*.

In either case, it will not be enough to resist the Supervenience Argument if for every action there is simply *some* agent who has performed an earlier choosy action. We can, following van Inwagen (1989), consider an agent-relativized version of the "N" operator, where "$\mathrm{N}_S p$," means "p and S does not have, and never has had, any choice about whether p." If every agent who acts choosily performs a first choosy act, then we can repeat the Supervenience Argument for each agent, relying on the agent-relativized "N." So the objection, if it is to have any bite, must be that at least some agents are such that either (a) for every time t, they perform a choosy action earlier than t, or (b) for some time t_0, they have never acted choosily up through t_0 but, for every time t after t_0, they have performed a (distinct) choosy act by t. Call agents of the first sort *eternal choosers*; call agents of the second sort, *dense choosers*.

Note first that it is extremely implausible that human beings are, or even could be, eternal or dense choosers. It is perhaps obvious that no human being has in fact existed through every moment of the universe's history. (Even if there is no first moment to our universe, our best science suggests there is plenty

of its history during which there were no humans.) Whether or not a being that had existed eternally could count as "human" is perhaps contentious; but even if we thought it were possible for there to be such eternal humans, we could consider "human*"s: humans who have not existed throughout every moment of a universe's history. It seems plausible that all actual humans are also necessarily humans*,[17] in which case none of *us* could be eternal choosers.

What of dense choosers? It seems plausible that there is a minimum temporal duration for human actions. That is, there should be some temporal length l such that we cannot perform any actions that take less time than l. We just can't act that fast, thanks to the way we're made. Likewise, it seems that we can only perform a finite number of actions at a time.[18] But these two considerations keep us from doing an infinite number of actions (choosy or not) in a finite stretch of time. So we can't be dense choosers.

Both of these arguments appeal to contingent facts about the kind of agents that populate our universe: we couldn't have been around for all of time; we can't react arbitrarily fast; and we can't do an infinite number of things at once. A foe of the Supervenience Argument might respond that this is irrelevant, insisting that the incompatibility of free will and naturalism should be understood as the impossibility of *any* free agents in *any* naturalistic universe. If I haven't shown that naturalistic eternal or dense choosers are impossible, I haven't shown that ($\beta\Box$) rules out the possibility of such a universe, so I haven't shown that ($\beta\Box$) rules out the compatibility of free will and naturalism.[19]

Warfield (2000, esp. §2) has argued for something similar with respect to the Consequence Argument. The Consequence Argument argues from contingent facts about us — that we have no choice about the laws, and that we have no choice about the remote past[20] — to a purported necessary conclusion. Warfield's response is to reject the Consequence Argument and place another argument in its stead.[21] And, as seems clear, any rejection of the Supervenience Argument along these lines would threaten the Consequence Argument as well.

However, I do not share Warfield's apparent conviction that the only incompatibility theses of interest are of the form "it is impossible that thesis T be true and there be free agents." Supposing, as I have argued, that *humans* cannot be eternal or dense choosers, and supposing, as I have not argued, that *humans* can-

[17] If we accept some sort of essentiality-of-origins thesis, this follows almost immediately.

[18] I am here speaking with the coarse-grained theorist. If the fine-grained theory is true, we do this all the time; but we will probably not be able to perform an infinite number of actions, each of which is in a different equivalence class of actions, where for some time t there is some action in each of these equivalence classes occurring at t.

[19] Thanks to Tom Crisp for pressing me to consider this reply.

[20] Which, if we were eternal choosers, we would have; thanks here to Joe Campbell.

[21] I do not consider his new argument here.

not have a choice about the past or laws, we can consider incompatibility theses of the form "it is impossible that T be true and there be free humans." Admittedly, this is weaker than the sort of claim Warfield prefers, but — given that we are all humans — it certainly seems like something we should be worried about.[22] And, insofar as we think the Consequence Argument gives us humans reason to worry about the incompatibility of human freedom and determinism, we should think the Supervenience Argument gives us reason to worry about the incompatibility of human freedom and naturalism.

Objection 2: In all discussions of r's supervenience base, you supposed that no event subvenient on r occurs after r. If a microphysical event g both subvened on and postdated r, though, then you would have no reason for thinking that NG is true (where G expresses the occurrence of g) and thus could not use ($\beta\Box$) to conclude that NR is true.

Before tackling this objection, consider another: perhaps there is backwards causation, and so someone has a choice about the truth of IN because they can do something after r which, if they did it, would cause *in* to not have occurred. This complaint needs little attention, though: in addition to undermining the Consequence Argument (since backwards causation makes it much more plausible that $\sim$NP is true), backwards causation simply seems, to most minds, too farfetched to serve in the defense of free will. (Even if it is possible, nobody wants to say that backwards causation is *required* for free will.)

On the other hand, backwards supervenience does not suffer from this *prima facie* implausibility. Some actions do seem to supervene on future events: if Jan shoots and kills Ron, then Jan's killing of Ron is arguably an action that is finished once her finger finishes pulling the trigger but that supervenes on the (future) event of Ron's death.

This response is only available to someone who holds the fine-grained theory of action-individuation. On a coarse-grained account, Jan's single action can be described in a number of ways: as a moving of her finger, as a shooting of a gun, or as a killing of Ron. If Ron had not died from the wound, then her action would not be accurately describable as a killing of Ron, but it still would have occurred. Thus, Jan's action does *not* supervene on a future event, although the fact that it can be described in a certain way does.

On a fine-grained account, Jan performed a number of distinct actions, including a moving of her finger, a shooting of a gun, and a killing of Ron. The

[22] If you do not accept that the relevant properties are essential to humans, consider a kind of humans* for which the relevant properties *are* essential, and ask yourself whether or not you are a human*. If you are, you should find the relevant incompatibility thesis worth worrying about.

last of these is plausibly thought of as supervening on Ron's death — it is not possible that Jan shoot Ron, Ron die the death he in fact died (the one involving the bullet that came out of the gun Jan fired), and Jan not kill Ron.

Call an event *e immediate* if and only if *e* does not supervene on any future events. As in the first objection, it will be useful to consider this as an objection to the version of the Supervenience Argument that makes use of an agent-relativized "N" operator. For an arbitrary agent *S*, call an event *e* such that $E \;\&\; \sim \mathrm{N}_S E$ is true *S-choosy*. I will argue that, for an arbitrary agent *S*, no event, immediate or non-immediate, is *S*-choosy. The argument proceeds from two claims: (a) some non-immediate events are *S*-choosy only if some immediate event is *S*-choosy, and (b) no immediate events are *S*-choosy. Since every event is either immediate or non-immediate, (a) and (b) together entail the non-existence of *S*-choosy events; since *S* was arbitrary, the non-existence of *S*-choosy events entails the non-existence of choosy events.

Here is the argument for (a). Let g be some non-immediate *S*-choosy event that occurs at a time t_1. Since g occurs at t_1, it is reasonable to assume that it has at least one immediate event in it supervenience base. (Otherwise, why say it occurred at t_1?) Call that event e_1. Since g is non-immediate, it must also supervene on some future event e_2, which we shall say occurs at t_2. If e_2 is in turn non-immediate, we can then consider concurrent and future events it supervenes on. Some of these future events may in turn be immediate or non-immediate, but if we are to avoid a regress, we will eventually come to some collection of immediate events in the future of a that form the post-t_1 portion of a supervenience base for g. (We can't have non-immediate events all the way down.) Since the number of future immediate events will not matter for the argument I am about to give, suppose that g supervenes on three immediate events: e_1, which occurs at t_1, e_2, which occurs at t_2, and e_3, which occurs at t_3.

If E_1, E_2, E_3, and G express the occurrence of e_1, e_2, e_3, and g, respectively, then $\Box((E_1 \;\&\; E_2 \;\&\; E_3) \rightarrow G)$ is true.[23] Since g is *S*-choosy, $\sim \mathrm{N}_S G$ is true. Thus, by $(\beta\Box_S)$, the relevant analogue of $(\beta\Box)$, $\sim \mathrm{N}_S(E_1 \;\&\; E_2 \;\&\; E_3)$ is also true. Since $E_1 \;\&\; E_2 \;\&\; E_3$ is true, there must be something *S* could have done that would have ensured its falsity. If one of e_1, e_2, or e_3 is choosy, then we have found an immediate choosy event and our conclusion is reached. So suppose that none of them are choosy. In this case, there is something that someone could have done (but did not do) that might ensure the falsity of some conjunct of $E_1 \;\&\; E_2 \;\&\; E_3$ and *would*, as a result, ensure the falsity of $E_1 \;\&\; E_2 \;\&\; E_3$.

[23] Perhaps we would be better off saying that $\Box((L \;\&\; E_1 \;\&\; E_2 \;\&\; E_3) \rightarrow G)$ is true. For simplicity I suppress the reference to the laws of nature; the argument is not affected by the modifications required by adding *L*.

By the broad past principle, nothing that S could have done after t_2 even *might* have ensured the falsity of E_1 or E_2. Thus, if e_3 is not choosy, there must have been something someone could have done that would have ensured the falsity of E_1 & E_2 & E_3 sometime before t_2. Let t be the time at which this act could have been done. By the sorts of considerations that underwrote (T3), we may assume that if there is something S could have done but did not do, then there is some (actual) S-choosy act that occurred no later than the merely possible act could have occurred. Call this act a, and suppose it occurred at t^*. Since t^* is no later than t, it is sometime earlier than t_2.

Since a is S-choosy, if it is immediate, then the claim we are trying to prove is true. If a is not immediate, we repeat the above argument replacing a and its supervenience base for g and its. If we *never* come to some immediate S-choosy act, then we face a regress, which can take either of two forms, one in which S must be an eternal chooser, and one in which S must be a dense chooser. Either options faces the problems that beset Objection 1 above.[24] If the objector wants to avoid these problems, she must concede that some immediate event is choosy.

The argument for (b) is simpler. Consider the first element of the class of immediate S-choosy events (or an arbitrary immediate event in the first choosy equivalence class of events). Note that, in the Supervenience Argument, in order to show that $\mathrm{N}_S R$ is true we only needed to show that true conjunctions of the laws of nature and propositions expressing the occurrence of microphysical events which causally contributed to r's supervenience base (or, as with x, were in this supervenience base) were such that S could not have ensured their falsity. But, while it may be plausible that some events are non-immediate, it is quite implausible (at least given naturalism) to hold that there is *any* event r that does not allow for some exhaustive list of immediate microphysical events causally contributing to its occurrence. Thus we shall be able to find the needed premises for an instance of the Supervenience Argument to show that r_1 is not choosy. In this case, there are no immediate S-choosy events, and thus (by claim (a)) there are also no non-immediate S-choosy events.

Objection 3: S had a choice about the occurrence of x. Thus, $\sim\mathrm{N}X$ is true.

I claim that x is not choosy. If my interlocutor simply *announces*, *sans* argument, that x is choosy after all, there is little I can offer by way of response beyond an incredulous stare. We generally do not think that objectively undetermined microphysical events are something that we have a choice about. When we

[24]Of course, here as there, these considerations support the Supervenience Argument insofar as we only use it to argue for the claim that *human* freedom is incompatible with naturalism. I assume for the balance of this paper that this is indeed the thesis we are worried about.

do think we have a choice about some undetermined microphysical event, it is generally because we think there is something we can do ahead of time to keep the event from occurring (e.g., destroying a radioactive sample to keep one of its atoms from decaying later; cf. Widerker 1987). Clearly, x is not choosy in *this* way. Why, then, should we think it has a special status not given to most undetermined microphysical events?

Imagine a fictional naturalistic libertarian, Nat, who responds as follows. "What we ought to have learned from Jaegwon Kim's exclusion arguments (e.g. 1998) and other work in mental causation is that we need to countenance top-down causation. The action r was caused by a belief-desire complex, db. Thus, db's supervenience base was at least partly causally responsible for r's supervenience base. But, in the sort of top-down view I advocate, this means that it is *also* true that db was partly causally responsible for r's supervenience base. (Granted, this means that r was overdetermined, but so what? I can learn to live with this sort of overdetermination (cf. Sider 2003).) In fact, I claim that one of the elements of r's supervenience base that db caused was x, although this causation was indeterministic. And since db — S's belief-desire complex — caused x, x is something that S had a choice about."[25]

I do not say that a naturalistic libertarian will have to respond exactly like Nat, but I suspect that any attempt to argue for the falsity of NX will involve specifying some special relation S bears towards x's occurrence — causal responsibility in the "right" sort of way, for instance. Any such argument will be irrelevant, though. I will happily grant that x was caused by S's mental states and even that S is, in some causal sense, "responsible" for x. Nonetheless, it does not appear that S is therefore in any better position to have a choice about whether x occurs.

Consider again McKay and Johnson's coin-tossing case. According to the intuitions driving that case, N(*the coin does not land heads*) is true, because the coin-tosser is unable to do anything to ensure that the coin lands heads. Nonetheless, if the coin-tosser actually *tosses* the coin, and it lands tails, it may very well be that her actions *caused*, and that she was in some way *causally responsible*, for the coin not landing heads.

I enjoin readers who feel tempted to say that S could have ensured the falsity of X to ask themselves whether they think that S could have done something to ensure the non-occurrence of a microphysical undetermined event *in the very same way* that a coin-tosser could *not* have done something to ensure that a coin lands tails. For my part, I cannot see any relevant difference between a tosser's abilities to ensure various outcomes of a coin toss and S's abilities to ensure

[25]Thanks here to Tom Crisp.

various outcomes of an undetermined event; and I suspect that I am not alone in this.

Of course, the foregoing hardly counts as an argument against the claim that S could have ensured the falsity of X. I have no such argument. Nor, I suspect, does my opponent have an argument *for* that claim. What we have instead is a disagreement about which claims are more implausible than which. I find the claim that individuals can have a non-derivative choice about the occurrence of objectively undetermined microphysical events — that is, that individuals can have a choice about such events that does not require them to have a choice about the occurrence of some prior events — more implausible than the claim that naturalistic agents cannot have the sort of freedom that libertarians want. And I have little to say that could convince any who find this ranking reversed.

This does not rob the Supervenience Argument of its bite, though. Just as I cannot convince my opponent that the truth of $\sim\mathrm{N}X$ is more implausible than the incompatibility of free will and naturalism, no libertarian will be able to convince a committed compatibilist that the incompatibility of free will and determinism is more plausible than the truth of $\sim\mathrm{N}(P \,\&\, L)$. Presumably, though, the libertarian does not feel threatened by this, since her goal is not to convert the compatibilist but rather to persuade the agnostic audience discussed in section 3. She is confident that most of that audience agrees with her about which claims are more plausible than which. In a similar vein, I feel confident that most of my agnostic audience — or, at least, most of those with naturalistic leanings — will agree with my assessment of the relative plausibility of the relevant claims here.

A naturalistic libertarian may agree that, insofar as S's choice about x is taken to be, in some important sense, "non-derivative," it is implausible that S had any choice about x. "But," he may say further, "S's choice *isn't* non-derivative after all. S has a choice about x because S has a choice about r. I claim that free will is compatible with naturalism, so I hold that S could have avoided r-ing. But if S had avoided r-ing, x would not have occurred; thus, S could have ensured the falsity of X. But S's X-falsifying powers are not in any sense 'non-derivative' — S falsifies X by falsifying R."[26]

I might respond that my interlocutor has begged the question. I have given an argument concluding that r is not choosy; he counters the argument by claiming that it is choosy after all. But I do not respond this way, and for two reasons. First, some may insist that I beg the question if I do not let my opponents suppose $\mathrm{N}R$ is false, for I then (they may say) unfairly shield my premises from any objections that do not presuppose my argument's conclusion. I do not wish to

[26] An argument like this was suggested to me by Kirk Ludwig.

embroil myself in sticky issues about question-begging and burdens of proof, so I will not make that charge.

Second, however, I am inclined to agree with my interlocutor that free will *is* compatible with naturalism. I may even agree that r is choosy; I am, at least, not convinced that it isn't. My claim was never that the argument's conclusion was *true*. It was rather that the argument's conclusion would be objectionable to any naturalistic libertarian but that there was no clear way to avoid it without opening the door to similar compatibilist refutations of the Consequence Argument.

What I am hesitant about is granting my opponent that, if $\sim\mathrm{N}R$ is true, then $\sim\mathrm{N}S$ is so as well. This is a substantial claim and needs to be argued for. Presumably, the argument runs something like this: there is nothing anyone could have done that even *might* have ensured the falsity of DB & C & IN & L. Thus, if $\sim\mathrm{N}(DB \,\&\, C \,\&\, IN \,\&\, L \,\&\, X)$ is true, then so is $\sim\mathrm{N}X$ for reasons canvassed above. However, by hypothesis, $\Box((DB \,\&\, C \,\&\, IN \,\&\, L \,\&\, X) \rightarrow R)$ is true. So, if $\sim\mathrm{N}R$ is true, then so is $\sim\mathrm{N}(DB \,\&\, C \,\&\, IN \,\&\, L \,\&\, X)$ and therefore $\sim\mathrm{N}X$ is as well.

In order to make this argument valid, even granting the move from the truth of $\sim\mathrm{N}(DB \,\&\, C \,\&\, IN \,\&\, L \,\&\, X)$ to that of $\sim\mathrm{N}X$, we have to grant the validity of ($\beta\Box$). In this context, such a gift ought not be made. I might be inclined to give my opponent $\sim\mathrm{N}R$, and I might be inclined to give him ($\beta\Box$), but I see no reason he should get both. *One* reason is that, were I to allow both together, I (along with him) would be forced into the implausible claim that we can have a choice about undetermined microphysical events. And he cannot soften the blow by saying, "ahh, but S had a choice about x only because he had a choice about r, and so there's nothing mysterious about his choice over microphysical undetermined events — he gets it from his choice over his own actions." If a tragic accident severs my spinal cord and I can no longer move my legs, we think the the *reason* for this is that certain of my nerves no longer work correctly and that I cannot make them work correctly. Indeed, if I can move my legs I can make my nerves work correctly by moving my legs; but if after the accident my nerves suddenly start working properly again, we cannot dispel the apparent mystery in this by pointing out that I can move my legs and that I make my nerves work properly by moving my legs. We are mystified by my miraculous leg movement 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. Likewise, it is mysterious how S is able to have a choice about r *just because* it is mysterious how S is able to have a choice about undetermined microphysical events; and without an account of the latter, we have no account of the former.

To grant ($\beta\Box$), in this instance and along with the truth of $\sim$NR, will be to grant that there is no account of either S's choice about x or his choice about r. I find this claim implausible; I find it less plausible, in fact, than the conjunction of ($\beta\Box$) and $\sim$NR. Thus, given the implausibility of $\sim$NX, I take the supposition that $\sim$NR is true to constitute, in effect, a counterexample to ($\beta\Box$). Clearly, however, in this case rejecting the Supervenience Argument by claiming that $\sim$NR is true directly threatens the Consequence Argument. Even if a naturalistic libertarian finds ($\beta\Box$) harder to reject than the claim that NX is true, there are still general dialectical problems to be faced, for it now begins to look like ($\beta\Box$) is not truly supported by intuitive plausibility but rather by incompatibilist dogma. It will be difficult to persuade an agnostic audience that inexplicable powers over undetermined microphysical events are a price worth paying for the intuitive benefits of ($\beta\Box$) and the naturalistic incompatibilism that goes with it.

References

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

Bricker, Phillip (2006). "The Relation Between General and Particular: Entailment vs. Supervenience." In Dean W. Zimmerman (ed.), *Oxford Studies in Metaphysics*, volume 2, 251–287. Oxford University Press.

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.

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

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.

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

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

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

— (1998). *Mind in a Physical World*. Cambridge, Mass.: MIT 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*. New York: Oxford University 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.

Sider, Theodore (2003). "What's So Bad About Overdetermination?" *Philosophy and Phenomenological Research* 67: 719–726.

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.

— (1989). "When Is the Will Free?" *Philosophical Perspectives* 3: 399–422.

— (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 for Incompatibilism." *Philosophical Perspectives* 14: 167–180.

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