17 November 2020

Free Will and the Cross-Level Consequence Argument

Jonathan Birch

Department of Philosophy, Logic and Scientific Method,

London School of Economics and Political Science,

Houghton Street, London, WC2A 2AE, UK.

j.birch2@lse.ac.uk

http://personal.lse.ac.uk/birchj1

Abstract

Christian List has recently constructed a novel formal framework for representing the relationship between free will and determinism. At its core is a distinction between physical and agential levels of description. List has argued that, since the consequence argument cannot be reconstructed within this framework, the consequence argument rests on a ‘category mistake’: an illicit conflation of the physical and agential levels. I show that an expanded version of List’s framework allows the construction of a cross-level consequence argument.

Christian List (2014, 2019a, b) has constructed a novel formal framework for representing the relationship between free will and determinism. The key innovation of the framework is that the physical and agential levels of description are explicitly represented, with each level having its own modal operators. List uses this framework to argue for a 'libertarian compatibilist' view on which the agential possibility of doing otherwise is compatible with determinism at the physical level. This is a version of compatibilism, in so far as it defends the compatibility of free will and physical determinism, but it also agrees with the libertarian that an agent 'could have done otherwise', provided 'could' is understood as a claim about agential, not physical, possibility.

This article constructs, within a slightly expanded version of List's framework, a valid cross-level variant of van Inwagen's (1983) consequence argument for the incompatibility of free will and determinism. In very informal terms, the consequence argument states that since we cannot change the laws of nature or the initial conditions of the universe, and since, given determinism, our actions now are logical consequences of these facts, we do not have free will if determinism is true. To avoid lengthy exposition, I will assume familiarity with van Inwagen's formalization of the consequence argument in terms of "Rule Beta" (i.e. the third formulation given in van Inwagen 1983, Chapter 3). List (2019a) has claimed that this argument cannot be reconstructed within his framework, because it would involve a "category mistake". The aim of this article is to show that such a reconstruction is possible and need involve no such mistake.

This cross-level consequence argument is of philosophical interest in its own right, since it has two important advantages over previous formulations: it explicitly incorporates the physical and agential levels of description, and it spells out the meaning of the relevant agent-level modal operator in a precise way. van Inwagen's formulation of the consequence argument notoriously relies on an informally characterized modal operator, *N*, that admits of multiple conflicting interpretations, spawning a large literature in which *N* is understood in various ways (e.g. Carlson 2000; Blum 2000; Huemer 2000; Beebee 2002; Campbell 2007; Huemer 2008; Pruss 2013; Gustafsson 2017). List's framework allows us to avoid van Inwagen's *N* and to formulate the argument using more precisely defined modal operators.

1. List's framework

List's (2014, 2019a) approach is based on models of dynamical systems with laws that are deterministic at the physical level and indeterministic at the agential level (see also List and Pivato 2015). This section explains the basic formal framework. I will follow the exposition of List (2019a), with one important difference, as explained below.

Let *S* denote the set of all possible physical states of the system, which are each fully specified and mutually exclusive. Let *T* denote the set of all points in time, where *T* is linearly ordered. A *physical history* is a function, denoted *h*, from *T* into *S*, which assigns to each point in time the corresponding state. Let Ω denote the set of all logically possible physical histories. Physical-level propositions are (extensionally) subsets of Ω, though we normally use sentences in a language to pick them out. A proposition *p* is *true* at some history *h* if and only if *h* is contained in the relevant subset of Ω.

To introduce physical-level modal operators, we define an accessibility relation between the elements of Ω. Whether one physical history is accessible from another depends on the time in question. We can posit, following List (2014, p. 164; 2019a, p. 261), that history h is *physically accessible* from history h' at time t if and only if the two histories have the same initial segment up to time t and diverge, at most, thereafter. A physical-level proposition p is *physically necessary* in history h at time t if and only if p is true in all histories h' physically accessible from h at t. Similarly, p is *physically possible* in history h at time t if and only if p is true in some history h' physically accessible from h at t.

To introduce agent-level propositions and modal operators, we need to re-describe the system. Let $\mathbb{S}$ denote the set of all logically possible states as described at the agential level. Each state in $\mathbb{S}$ specifies the mental attitudes and actions of all agents in the system at the time in question. We assume that the agential states in $\mathbb{S}$ supervene on the physical states in S, meaning that there exists a function σ from S into $\mathbb{S}$ in which multiple physical states may be mapped on to the same agential state. Like physical states, different agential states are mutually exclusive.

An *agential history* is a temporal path of the system through its agent-level state space. Formally, this is a function $\boldsymbol{h}$ from T into $\mathbb{S}$. Each physical history h gives rise to a corresponding agential history $\boldsymbol{h}$, obtained by applying the supervenience mapping σ to the given physical history. Formally, we write $\boldsymbol{h} = \sigma(h)$. Let $\boldsymbol{\Omega}$ denote the set of all logically possible agential histories. An agential-level proposition, then, is (extensionally) a subset of $\boldsymbol{\Omega}$.

Let us now define necessity and possibility at the agential level. This is where I must take issue with List's exposition. List defines agential accessibility as follows: an agential history $\boldsymbol{h}$ is *agentially accessible* from another such history $\boldsymbol{h}'$ at time t if and only if the two histories have the same initial segment up to time t and diverge, at most, thereafter (List 2014, p. 165; List 2019a, p. 262). However, it begs the question against an incompatibilist opponent to assume, as a matter of definition, that only the *agential* past at t, and not the physical past, constrains agential accessibility. Since my aim in this paper is to construct an argument against compatibilism, I want to avoid any such presupposition. Accordingly, I will take the agential accessibility of one agential history from another at t as a primitive relation.

We can now define agential necessity and possibility with reference to agential accessibility. A purely agent-level proposition $\boldsymbol{p}$ is *agentially necessary* in agential history $\boldsymbol{h}$ at time t if and only if $\boldsymbol{p}$ is true in all agential histories $\boldsymbol{h}'$ agentially accessible from $\boldsymbol{h}$ at t. Similarly, $\boldsymbol{p}$ is *agentially possible* in agential history $\boldsymbol{h}$ at time t if and only if $\boldsymbol{p}$ is true in some history $\boldsymbol{h}'$ agentially accessible from $\boldsymbol{h}$ at t.

In List's framework, the physical level past, plus a set of deterministic physical laws, determine the physical-level future. Meanwhile, the agent-level past, plus a set of deterministic laws of agency (laws relating the agential facts at one time to the agential facts a moment later), *would* determine the agent-level future—if there were any such laws. In reality, however, the laws of the agent-level are indeterministic: the agent-level facts at a given time leave various futures open. In light of this, and given List's definition of agential accessibility, alternative possibilities are agentially possible at any time. This is not threatened by physical-level determinism, as long as the various physical histories that are compatible with the agential past up to that time lead to a range different agential futures.

List sees this as an argument for compatibilism. However, since his definition of agential accessibility presupposes compatibilism, it can't be used in a non-question-begging argument for it. It is more accurate to say the framework provides a formal representation of compatibilism. By contrast, if we take agential accessibility as a primitive relation (as urged above), the framework leaves open the question of whether the physical-level past plus a deterministic set of physical laws are compatible with the agential possibility of alternative actions. This gives us a framework within which an argument against compatibilism can be constructed.

2. An expansion of the framework

List (2019a, pp. 266-268) argues that there is no way to formulate a compelling version of the consequence argument within his framework. We can, he argues, formulate a version of the argument using only the physical-level modal operators, but this says nothing about the agent level. We can also formulate a version using only the agent-level modal operators, but, since agent-level determinism is clearly false (i.e., actions are *not* determined by past actions plus laws of agency), this version is clearly unsound. What List argues we cannot do is move, as van Inwagen does, between claims about the physical level and claims about the agent level. To do this, List argues, is to make a 'category mistake'. List therefore concludes that his version of compatibilism evades van Inwagen's challenge.

However, it is possible to formulate, within a slightly expanded version of List's formalism, a 'cross-level' version of the consequence argument that specifically targets his version of compatibilism. To formulate such an argument, we first need a way of expressing claims about the relations between the physical and agential levels. We can do this by expanding List's framework in a straightforward way.

Consider the Cartesian product $\Omega\times\boldsymbol{\Omega}$, a set comprising *all logically possible pairs of physical and agential histories*. It will be convenient to refer to these pairs of histories simply as "worlds" from now on. If it is logically possible that the supervenience mapping σ fails to obtain, $\Omega\times\boldsymbol{\Omega}$ will contain worlds that violate this mapping. Moreover, if the supervenience mapping is metaphysically but not logically necessary, $\Omega\times\boldsymbol{\Omega}$ will contain some metaphysically impossible worlds. It will be useful, given this, to have a way of denoting the subset of $\Omega\times\boldsymbol{\Omega}$ comprising all and only those worlds $(h, \boldsymbol{h})$ such that $\boldsymbol{h} = \sigma(h)$. This is the set of all logically possible worlds compatible with the supervenience mapping $\sigma(h)$. Call this subset Ψ. Ψ is the proposition that the agential level supervenes on the physical level in accordance with the supervenience mapping $\sigma(h)$.

List can have no objection to this expansion of the framework, because such an expansion is needed in order to assert Ψ, i.e. to assert the proposition that the agent-level supervenes on the physical-level in accordance with $\sigma(h)$. If List refused to allow this expansion, but continued to maintain that cross-level propositions which cannot be expressed within the framework are category mistakes, he would be forced to conclude that any assertion of the supervenience of one level on another is also a category mistake. But the existence of such supervenience relations is a foundational assumption of List's approach.

Physical-level propositions and agent-level propositions that are actually true will both map to subsets of Ψ. Moreover, there will be subsets of Ψ that are most naturally picked out by sentences that mix physical and agential language, such as the sentence that there is an agent in some region only if there is physical matter in that region, or the sentence that a certain action requires a certain amount of energy to perform.

3. Cross-level bridging principles

We are now in a position to say more about the relation between the agential and physical levels. In particular, we can introduce a bridging principle:

> *Bridging Principle 1:* A world (h', $\boldsymbol{h}'$) is agentially accessible from another world (h, $\boldsymbol{h}$) at t if and only if $\boldsymbol{h}'$ is agentially accessible from $\boldsymbol{h}$ at t and h' is physically accessible from h at t.

From this agential accessibility relation between worlds, we can derive an expanded notion of agential possibility at a world, where p denotes *any* proposition within $\Omega\times\boldsymbol{\Omega}$, including physical, agential, and mixed-level propositions:

> *Bridging Principle 2:* A proposition p is agentially possible in (h, $\boldsymbol{h}$) at t if and only if there is a world (h', $\boldsymbol{h}'$) at which p is true that is agentially accessible from (h, $\boldsymbol{h}$) at t.

Informally, these bridging principles expand the notions of agential accessibility and agential possibility, introducing a sense in which propositions that are not themselves purely agent-level can be agentially possible at a world. If we endorse the bridging principles then we can talk meaningfully of the agential possibility of any proposition, regardless of whether it is an agent-level proposition, a physical-level proposition, or a mixed-level proposition.

It might be objected that the bridging principles beg the question against a compatibilist by tying agential possibility to physical possibility, so that a proposition that is physically impossible at *t* cannot be agentially possible at *t*. However, I see the dialectical situation like this: true enough, one way to be a compatibilist is to deny the above bridging principles. But many compatibilists, I take it, will not want to resort to denying the bridging principles, because they accept that physically impossible propositions are not agentially possible. Rather, they will want to show that the bridging principles pose no threat to compatibilism. For these compatibilists, the question arises: can one consistently accept the bridging principles and still be a compatibilist? This is the question that is at issue in the following discussion.

I suspect List would regard this expansion of the framework as a “dramatic redefinition of the semantics of the agential-level modal operators” (List 2019a, p. 271). Of such redefinitions, he comments that “the agential ‘can’ is a higher-level notion; it is not to be found at the fine-grained level at which any such redefinition would attempt to relocate it” (p. 271). The concern appears to be that, by constructing an expanded agential possibility operator that may take physical-level propositions as inputs, we are severing it from the ordinary meaning of the word “can”.

I have two responses to this. First, note that many ordinary statements apply “can” to physical-level events without loss of meaning. For example, I might say “I can try to stop a virus spreading, but I cannot stop a virus mutating”. On the face of it, this is a true statement, not a category mistake, even though the mutation of a virus is a physical-level event. Second, I reject the underlying assumption that our intuitions about ordinary language should constrain the construction of formal frameworks for debating free will. Better, I suggest, to see where the bridging principles lead, while allowing that one possible compatibilist escape route is simply to deny that the bridging principles capture a legitimate sense of agential possibility.

4. Cross-level modal operators

Now let us define an operator $\Box$ such that $\Box\, p$ is true if and only if p (which may be an agent-level proposition, a physical-level proposition, or a cross-level proposition) is true at all worlds $(h, \boldsymbol{h}) \in \Omega\times\boldsymbol{\Omega}$ and at all times t. Thus defined, the $\Box$ operator is implied by and close to logical necessity, although some propositions may be true at all logically possible pairs of physical and agential histories without being strictly logically necessary (e.g. the proposition that there is an agential level). It implies but is not implied by metaphysical necessity, because $\Omega\times\boldsymbol{\Omega}$ may contain worlds at which metaphysically necessary truths are violated. As noted above, this will be the case if the supervenience of the agential level on the physical level is metaphysically but not logically necessary. Similarly, it implies but is not implied by nomological necessity, because $\Omega\times\boldsymbol{\Omega}$ contains worlds at which the actual laws of nature are violated. The $\Box$

operator allows us to express the idea that the initial state of the universe, the laws of physics, and the supervenience mapping between levels determine the agent-level facts.

We now need to introduce a variant of van Inwagen's *N*, defined in terms of List's formalism. Let us call this **N**:

> $\mathbf{N}p$ in world $(h, \boldsymbol{h})$ at time t if and only if (by definition) not-p is not agentially possible in $(h, \boldsymbol{h})$ at t.

If we grant that there is no category mistake involved in the expanded notion of agential possibility introduced above, then there is also no category mistake involved in formulating **N**. A difference with van Inwagen's *N* is that our **N** only yields an output if a world $(h, \boldsymbol{h})$ and a time point t are both specified. If no world or no time point is specified, the agential possibility operator yields no output. $\mathbf{N}p$ informally means that there is no agentially possible future at $(h, \boldsymbol{h})$ and t in which p is false.

5. A cross-level consequence argument

With these pieces in place, we can now reconstruct the consequence argument. We need the following two inference rules:

> *Rule Alpha:* If $\Box p$, infer: $\mathbf{N}p$ at all $(h, \boldsymbol{h})$, t
>
> *Rule Beta:* If $\mathbf{N}p$ and $\mathbf{N}(p \rightarrow q)$ at $(h, \boldsymbol{h})$, t, infer: $\mathbf{N}q$ at $(h, \boldsymbol{h})$, t.

These rules are supported by our expanded definition of agential possibility and by our bridging principles. Rule Alpha is valid because, if p is true at all worlds, then, trivially, there can be no world at which p is false that is agentially accessible from (h, $\boldsymbol{h}$) at t. Rule Beta is valid because, if p is true at all worlds accessible from (h, $\boldsymbol{h}$) at t, and if ($p \rightarrow q$) is also true at all these worlds, then none of these are worlds at which q is false.

We also need the following notation: p_0 is a physical proposition uniquely specifying the initial physical state of the focal agent's universe; $\boldsymbol{p_1}$ is an agent-level proposition uniquely specifying the agential state of the focal agent at some later time t_1 (e.g. I raised my right-hand at t); l is a proposition specifying the laws of the physical level in the focal agent's universe; and Ψ (as introduced above) is the proposition that a specific supervenience mapping $\sigma(h)$ from physical histories to agential histories obtains.

Note that, assuming determinism, the conjunction of the initial physical state of the universe and the laws, p_0 & l, can be identified with a unique physical-level history h_1 at which both conjuncts are true. Note further that this physical-level history maps, by the supervenience mapping $\sigma(h)$, to a unique agent-level history, $\boldsymbol{h_1}$, which in turn determines $\boldsymbol{p_1}$, the state of the focal agent at t_1. Thus, assuming determinism and the supervenience of the agential on the physical, the conjunction p_0 & l & Ψ logically entails $\boldsymbol{p_1}$, and thus $\Box$ ((p_0 & l & Ψ) $\rightarrow \boldsymbol{p_1}$) is true.

We can use this observation to run a cross-level consequence argument from determinism and supervenience to the impossibility of doing otherwise, as follows.

Cross-level consequence argument:

1. $\Box\, ((p_0 \,\&\, l \,\&\, \Psi) \rightarrow \boldsymbol{p_1})$ (Determinism plus supervenience)
2. $\Box\, (p_0 \rightarrow (l \rightarrow (\Psi \rightarrow \boldsymbol{p_1})))$ (Rearrangement of 1)
3. $\mathbf{N}\, (p_0 \rightarrow (l \rightarrow (\Psi \rightarrow \boldsymbol{p_1})))$ at $(h_1, \boldsymbol{h_1}), t_1$ (from Rule Alpha)
4. $\mathbf{N}\, p_0$ at $(h_1, \boldsymbol{h_1}), t_1$ (Fixed Past)
5. $\mathbf{N}\, (l \rightarrow (\Psi \rightarrow \boldsymbol{p}))$ at $(h_1, \boldsymbol{h_1}), t_1$ (from Rule Beta)
6. $\mathbf{N}\, l$ at $(h_1, \boldsymbol{h_1}), t_1$ (Fixed Laws)
7. $\mathbf{N}\, (\Psi \rightarrow \boldsymbol{p_1})$ at $(h_1, \boldsymbol{h_1}), t_1$ (from Rule Beta)
8. $\mathbf{N}\, \Psi$ at $(h_1, \boldsymbol{h_1}), t_1$ (Fixed Mapping)
9. $\mathbf{N}\, \boldsymbol{p_1}$ at $(h_1, \boldsymbol{h_1}), t_1$ (from Rule Beta)

This is a variant of van Inwagen's consequence argument. It differs from van Inwagen's original formulation in two respects. First, rather than relying on an informally characterized modal operator *N*, the work is done by the operator **N** which is defined in terms of an agential accessibility relation. Second, rather than implicitly assuming the metaphysical determination of agent-level facts by physical-level facts, the above formulation *explicitly* incorporates this determination via the inclusion of the proposition Ψ that the supervenience mapping $\sigma(h)$ obtains, along with an extra premise, Fixed Mapping, asserting that histories where the supervenience mapping fails to obtain are not agentially possible.

The argument as a whole proceeds from determinism and supervenience to the agential impossibility of doing otherwise, presenting an obstacle to List's 'libertarian compatibilism'. It poses a greater threat than either of the single-level consequence arguments considered by List (2019a, pp. 266-268). Unlike List's purely physical-level version, it establishes the agential necessity of an agent-level proposition. And unlike List's purely agent-level version, it does not rely on the (implausible) assumption that the agent-level past and the agent-level laws determine the agent-level present.

References

Beebee, H. (2002). Reply to Huemer on the consequence argument. *Philosophical Review* 111:235-241.

Blum, A. (2000). 'N'. *Analysis* 60:284-286.

Brueckner, A. (2008). Retooling the consequence argument. *Analysis* 68:10–13.

Campbell, J. K. (2007). Free will and the necessity of the past. *Analysis* 67:105-111.

Carlson, E. (2000). Incompatibilism and the transfer of power necessity. *Noûs* 34:277-290.

Carlson, E. 2003. Counterexamples to Principle Beta: a response to Crisp and Warfield. *Philosophy and Phenomenological Research* 66: 730–37.

Gustafsson, J. E. 2017. A strengthening of the Consequence Argument for incompatibilism. *Analysis* 77:705-715.

Lewis, D. K. 1981. Are we free to break the laws? *Theoria* 47: 113–21.

List, C. 2014. Free will, determinism, and the possibility of doing otherwise. *Noûs* 48:156-178.

List. C. and M. Pivato. 2015. Emergent chance. *Philosophical Review* 124:119-152.

List, C. 2019a. What's wrong with the Consequence Argument: A libertarian compatibilist response. *Proceedings of the Aristotelian Society* 119:253-274.

List, C. 2019b. *Why Free Will is Real*. Cambridge, MA: Harvard University Press.

McKay, T. J. and D. Johnson. 1996. A reconsideration of an argument against compatibilism. *Philosophical Topics* 24:113–22.

Pruss, A. R. (2013). Incompatibilism proved. *Canadian Journal of Philosophy* 43:430-437.

van Inwagen, P. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.

Huemer, M. (2000). Van Inwagen’s Consequence Argument. *Philosophical Review* 109:525-544.