INCOMPATIBILISM PROVED

ALEXANDER R. PRUSS

ABSTRACT. The consequence argument attempts to show that incompatibilism is true by showing that if there is determinism, then we never had, have or will have any choice about anything. Much of the debate on the consequence argument has focused on the "beta" transfer principle, and its improvements. We shall show that on an appropriate definition of "never have had, have or will have any choice", a version of the beta principle is a *theorem* given one plausible axiom for counterfactuals (weakening). Instead of being about transfer principles, the debate should be over whether the distant past and laws are up to us.

1. BACKGROUND

Consequence arguments attempt to establish that if determinism is true, no one has a choice about anything. More precisely, let L be a conjunction of all the laws and let P be a statement of the complete state of the world in the pre-human past. For a statement r, let Nr be the statement that r holds and that no one ever had, has or will have a choice whether r holds.

Van Inwagen's (1983) consequence argument can be formulated as follows. Assume r is a true statement about something during human times, say that Jones is mowing the lawn at t_3, and assume that determinism holds. We need two rules of inference for the proof. The first is that we can derive no-choice-about from necessity:

ALPHA. $\Box q \vdash Nq$

The second rule is that no-choice-about transfers across material conditionals there is no choice about:

BETA. $Nq, N(q \supset r) \vdash Nr$

Given these rules, we can prove that no one ever had, has or will have a choice whether r holds, assuming that (a) the laws and the pre-human past entail r (determinism), (b) there is no choice about the laws, and (c) there is no choice about the pre-human past. The proofs in this paper will be given in a simple Fitch-style natural deduction system with a number of convenience rules including the catch-all "taut con" for tautological (i.e., truthfunctional) consequence. On the modal side, we will use three rules which yield a subset of system T, namely distribution, necessitation of tautology (T has necessitation of all theorems) and rule T:

DIST. $\Box(p \supset q) \vdash \Box p \supset \Box q$

NEC TAUT. If p is a tautology, then $\vdash \Box p$

T. $\Box p \vdash p$

Then the proof that we have no choice about p is:

1	$\Box(L \wedge P \supset r)$	
2	NL	
3	NP	
4	$\Box((L \wedge P \supset r) \supset (L \supset (P \supset r))$	nec taut
5	$\Box(L \wedge P \supset r) \supset \Box(L \supset (P \supset r))$	distribution, 4
6	$\Box(L \supset (P \supset r))$	$\supset$-elim, 1, 5
7	$N(L \supset (P \supset r))$	alpha, 6
8	$N(P \supset r)$	beta, 2, 6
9	Nr	beta, 3, 8

Much of the discussion in the literature has focused on beta, which was conclusively counterexampled by McKay and Johnson (1996), as van Inwagen (2000, 8) himself has admitted. The problem is that alpha, which is generally accepted in the literature, when conjoined with beta, yields the agglomeration rule:

AGG. $Np, Nq \vdash N(p \wedge q)$

For:

1	Np	
2	Nq	
3	$\Box[p \supset (q \supset (p \wedge q))]$	nec taut
4	$N[p \supset (q \supset (p \wedge q))]$	alpha, 3
5	$N(q \supset (p \wedge q))$	beta, 1, 4
6	$N(p \wedge q)$	beta, 2, 5

But agglomeration is invalid. Suppose you actually won't toss an indeterministic coin but can. Let p be the proposition that the coin won't land heads. Let q be the proposition that the coin won't land tails. Then Np, since p is true and you have no choice about p, because there is nothing you could do to make p false—you can't make the coin land heads. Similarly, Nq. On the other hand $N(p \wedge q)$ is false, because you do have a choice about $p \wedge q$—if you toss the coin, the conjunction $p \wedge q$ will be false, since the coin will land either heads or tails.

Widerker (1987) and Finch and Warfield (1998) have shown how one can replace beta by another plausible principle, which following Finch and Warfield I will refer to as beta-2, and which is not subject to the above counterexample:

BETA-2. $Np, \Box(p \supset q) \vdash Nq$

To run a consequence argument on this basis, these authors replace NL and NP in the assumptions of the proof by the single claim $N(P \wedge L)$. Without agglomeration, this single claim cannot be derived from NL and NP, but it is nonetheless plausible that we have no choice about $P \wedge L$. Then one doesn't even need alpha and the proof is a trivial consequence of beta-2:

1	$\Box(L \wedge P \supset r)$	
2	$N(L \wedge P)$	
3	Nr	beta-2, 1, 2

Unfortunately, all that Finch and Warfield (1998) say in favor of beta-2 is that it appears intuitively correct and that they cannot think of a counterexample while Widerker (1987) derives the rule from a complicated and controversial rule about prevention. But one can do rather better. We shall give a more rigorous definition of N (following the second suggestion in Huemer 2000, 529), and show that rule beta-2 follows from the plausible weakening rule for subjunctive conditionals, which rule is guaranteed to hold in Lewisian semantics. We shall also give a consequence argument that uses an alternate M operator.

2. Proving beta-2

To defend a version of the consequence argument, all that is needed (cf. Huemer 2000, 529–530) is to find *some* interpretation of Np such that

(a) the premises of the argument are true,
(b) the claim that Nr holds for all truths r about human times is incompatible with any humans having free will and
(c) beta-2 is valid.

If the interpretation sounds natural and fits with what van Inwagen says, that's all the better, however.

It is natural to read the claim that there was, is and will be no choice about r as saying that there is nothing that anyone can (ever) do that *would* falsify r:

N-DEF. Nr if and only if $r \wedge \sim \exists x \exists \alpha[\mathrm{Can}(x, \alpha) \wedge (\mathrm{Does}(x, \alpha) \,\Box\!\!\rightarrow \sim r)]$

where $\Box\!\!\rightarrow$ is a subjunctive conditional, x ranges over humans and α ranges over all possible past, present and future action types[1], such as *mowing the lawn at* t_3 (this is what Huemer [2000, 529] calls the "counterfactual sufficiency" interpretation). Take this to be the definition of Nr. Of course, analyzing $\mathrm{Can}(x, \alpha)$ would be a challenging task, but it will turn out that we will not need any controversial analysis—our proof works regardless of how we analyze it.

Only one premise of the beta-2 version of the consequence argument involves the N operator: $N(L \wedge P)$. And this premise remains plausible on this interpretation: it is implausible that there is something we can do such that were we to do it, the conjunction of laws and pre-human past wouldn't hold. The other premise was a standard definition of determinism, namely $\Box(L \wedge P \supset r)$ when r is about human times, and that is unaffected by our definition of N. Thus we have reason to think we have desideratum (a): the truth of the premises. Desideratum (b) is also very plausible: if Nr holds for all propositions r about human times, then plausibly nobody has free will. What remains is (c), the validity of rule beta-2.

To derive beta-2, we will use the rule of weakening for subjunctive conditionals:

[1]Or, to formulate this without typed variables, we can simply stipulate that $\mathrm{Can}(x, \alpha)$ is only true when x is a human and α a possible action type.

WEAKEN. $p \mathrel{\Box\!\!\rightarrow} q, \Box(q \supset r) \vdash p \mathrel{\Box\!\!\rightarrow} r$

In Lewis' semantics, weakening is easy to show. If p is impossible, then $p \mathrel{\Box\!\!\rightarrow} r$ holds trivially. If p is possible, then $p \mathrel{\Box\!\!\rightarrow} q$ holds if and only if there is a world w_1 at which p and q hold and which is closer to the actual world than any world at which p and $\sim q$ hold. Let w_1 be such a world. Since q entails r, we must also have r holding at w_1. Let w_2 now be any world at which p and $\sim r$ hold. Because q entails r, this is a world at which p and $\sim q$ hold and hence this world is further from the actual world than w_1 is, by choice of w_1. Thus, $p \mathrel{\Box\!\!\rightarrow} r$ holds.

Now we show that beta-2 can be derived by weakening given our definition of N.

1	Np	
2	$\Box(p \supset q)$	
3	$p \wedge \sim \exists x \exists \alpha[\mathrm{Can}(x,\alpha) \wedge (\mathrm{Does}(x,\alpha) \mathrel{\Box\!\!\rightarrow} \sim p)]$	def, 1
4	$\sim \exists x \exists \alpha[\mathrm{Can}(x,\alpha) \wedge (\mathrm{Does}(x,\alpha) \mathrel{\Box\!\!\rightarrow} \sim p)]$	$\wedge$-elim, 3
5	$\forall x \forall \alpha \sim[\mathrm{Can}(x,\alpha) \wedge (\mathrm{Does}(x,\alpha) \mathrel{\Box\!\!\rightarrow} \sim p)]$	De Morgan, 4
6	ba $\sim[\mathrm{Can}(b,a) \wedge (\mathrm{Does}(b,a) \mathrel{\Box\!\!\rightarrow} \sim p)]$	$\forall$-elim, 5
7	$\mathrm{Can}(b,a) \supset \sim(\mathrm{Does}(b,a) \mathrel{\Box\!\!\rightarrow} \sim p)$	taut con, 6
8	$\Box[(p \supset q) \supset (\sim q \supset \sim p)]$	nec taut
9	$\Box(p \supset q) \supset \Box(\sim q \supset \sim p)$	distr, 8
10	$\Box(\sim q \supset \sim p)$	$\supset$-elim, 2, 9
11	$\mathrm{Can}(b,a)$	
12	$\mathrm{Does}(b,a) \mathrel{\Box\!\!\rightarrow} \sim q$	
13	$\mathrm{Does}(b,a) \mathrel{\Box\!\!\rightarrow} \sim p$	weaken, 10, 12
14	$\sim(\mathrm{Does}(b,a) \mathrel{\Box\!\!\rightarrow} \sim p)$	$\supset$-elim, 7, 11
15	$\sim(\mathrm{Does}(b,a) \mathrel{\Box\!\!\rightarrow} \sim q)$	reductio, 12–14
16	$\mathrm{Can}(b,a) \supset \sim(\mathrm{Does}(b,a) \mathrel{\Box\!\!\rightarrow} \sim q)$	$\supset$-intro, 11–15
17	$\sim[\mathrm{Can}(b,a) \wedge (\mathrm{Does}(b,a) \mathrel{\Box\!\!\rightarrow} \sim q)]$	taut con, 16
18	$\forall x \forall \alpha \sim[\mathrm{Can}(x,\alpha) \wedge (\mathrm{Does}(x,\alpha) \mathrel{\Box\!\!\rightarrow} \sim q)]$	$\forall$-intro, 17
19	$\sim \exists x \exists \alpha[\mathrm{Can}(x,\alpha) \wedge (\mathrm{Does}(x,\alpha) \mathrel{\Box\!\!\rightarrow} \sim q)]$	De Morgan, 18
20	$p \supset q$	T, 2
21	p	$\wedge$-elim, 3
22	q	$\supset$-elim, 20, 21
23	$q \wedge \sim \exists x \exists \alpha[\mathrm{Can}(x,\alpha) \wedge (\mathrm{Does}(x,\alpha) \mathrel{\Box\!\!\rightarrow} \sim q)]$	$\wedge$-intro, 19, 22
24	Nq	def, 23

Combining this with the simple beta-2 based consequence argument, we see that given an uncontroversial modal logic weaker than system T, our counterfactual definition of N and the weakening rule for subjunctive conditionals, we can prove that we have no choice about a proposition r from the deterministic assumption that r is entailed by the conjunction of the laws and pre-human past and the assumption that we have no choice about that conjunction.

Could there be a counterexample to beta-2 on this reading of the N-operator? Perhaps. But it would be a substantive philosophical achievement, since it would also be a counterexample to weakening and hence to Lewis's semantics for counterfactuals, as well as to various modified versions of that semantics on which weakening still holds (e.g., Pruss 2007 and 2011).

The most promising strategy to criticize weakening seems to be to note that weakening implies that $p \mathrel{\Box\!\!\rightarrow} r$ for any r if p is impossible (this follows from the uncontroversial axiom $p \mathrel{\Box\!\!\rightarrow} p$ together with the fact that an impossibility entails everything). Defenders of non-trivial truth values for counterpossibles will say that $p \mathrel{\Box\!\!\rightarrow} r$ can fail to be true even if p is impossible.

However, a reasonable response is that the regular counterfactual conditional is different from the counterpossible conditional. Thus, while weakening does not hold for the counterpossible conditional, it does hold for the regular counterfactual conditional. And in fact it is very plausible that any counterexamples to weakening will involve an impossible antecedent. It is only in an *impossible* counterfactual scenario that we can't move from the claim that q would hold in that scenario to the claim that r would hold in the scenario when q entails r.

3. An alternate version

On our definition, to claim that Nr is to claim that r and that there is nothing we can do that *would* falsify this. Perhaps a compatibilist could claim that this isn't enough to make it correct to say that we have no choice about r. An alternate no-choice claim would be that no matter what action any humans did within their abilities, r would (still) be true:

M-def. Mr if and only if $r \wedge \forall x \forall \alpha [\mathrm{Can}(x,\alpha) \supset (\mathrm{Does}(x,\alpha) \mathrel{\Box\!\!\rightarrow} r)]$

This M operator is equivalent to Huemer's (2000) N_S operator. Also, it is left as an exercise to the reader to verify that by the Lewisian duality $(p \mathrel{\Box\!\!\rightarrow} {\sim} q) \Leftrightarrow {\sim}(p \mathrel{\Diamond\!\!\rightarrow} q)$, where $\mathrel{\Diamond\!\!\rightarrow}$ is the might-conditional, Mr is equivalent to what we get if we replace $\mathrel{\Box\!\!\rightarrow}$ with $\mathrel{\Diamond\!\!\rightarrow}$ in the definition of Nr.

Notice that the M-operator is agglomerative if this plausible conjunction rule holds for subjunctives:

conj. $(p \mathrel{\Box\!\!\rightarrow} r), (p \mathrel{\Box\!\!\rightarrow} s) \vdash (p \mathrel{\Box\!\!\rightarrow} r \wedge s)$

It is left to the interested reader to verify that given Mp and Mq, by the conjunction rule one can prove $M(p \wedge q)$. Moreover, the agglomerativeness of M would not be a problem. Take the McKay and Johnson (1996) counterexample to agglomerativeness, where I won't flip a coin, p says that the coin won't land heads and q says that the coin won't land heads. Then as in their counterexample, $M(p \wedge q)$ is false. But even if this false claim follows from Mp and Mq, that is not a problem since Mp is not true (and Mq is also not true), since there is an action α that I can take such that ${\sim}(\mathrm{Does}(\mathrm{I},\alpha) \mathrel{\Box\!\!\rightarrow} p)$, namely I can flip a coin. For it is false that were I to flip the coin, the coin wouldn't land heads—it might well (and it likewise might land tails)!

Note, however, that this gvies us reason to think that the M-operator is not what van Inwagen originally intended in his consequence argument, since van Inwagen denies agglomeration.

To run the simple Widerker-style argument on the basis of M, we need an analogue to beta-2:

GAMMA-2. $Mp, \Box(p \supset q) \vdash Mq$

For then we can argue:

1	$\Box(L \wedge P \supset r)$	
2	$M(L \wedge P)$	
3	Mr	gamma-2, 1, 2

And if we allow the conjunction rule, then we have agglomeration, and we can even replace $M(L \wedge P)$ with ML and MP.

The advantage of this version of the argument is that the conclusion Mr is more damaging to the compatibilist than the conclusion Nr: not only there is nothing we can do that *would* falsify r, but no matter what we did from among the things we can do, r would still be true, or by the would-might duality, there is nothing we can do that *might* falsify r. The downside of this argument is that $M(L \wedge P)$ may be a bit more controversial than $N(L \wedge P)$ (Beebee 2002 criticizes Huemer precisely for relying on, in our notation, ML).

What remains is to prove gamma-2. The proof is actually simpler than our proof of beta-2. It uses weakening but not the conjunction rule for subjunctives, and the only rule of modal logic that it uses is T:

1	Mp	
2	$\Box(p \supset q)$	
3	$p \wedge \forall x \forall \alpha[\mathrm{Can}(x,\alpha) \supset (\mathrm{Does}(x,\alpha) \mathbin{\Box\!\!\rightarrow} p)]$	def, 1
4	$\forall x \forall \alpha[\mathrm{Can}(x,\alpha) \supset (\mathrm{Does}(x,\alpha) \mathbin{\Box\!\!\rightarrow} p)]$	$\wedge$-elim, 3
5	ba | $\mathrm{Can}(b,a) \supset (\mathrm{Does}(b,a) \mathbin{\Box\!\!\rightarrow} p)$	$\forall$-elim, 4
6	| | $\mathrm{Can}(b,a)$	
7	| | $\mathrm{Does}(b,a) \mathbin{\Box\!\!\rightarrow} p$	$\supset$-elim, 5, 6
8	| | $\mathrm{Does}(b,a) \mathbin{\Box\!\!\rightarrow} q$	weaken, 2, 7
9	| $\mathrm{Can}(b,a) \supset (\mathrm{Does}(b,a) \mathbin{\Box\!\!\rightarrow} q)$	$\supset$-intro, 6–8
10	$\forall x \forall \alpha[\mathrm{Can}(x,\alpha) \supset (\mathrm{Does}(x,\alpha) \mathbin{\Box\!\!\rightarrow} q)]$	$\forall$-intro, 9
11	$p \supset q$	T, 2
12	p	$\wedge$-elim, 3
13	q	$\supset$-elim, 11, 12
14	$q \wedge \forall x \forall \alpha[\mathrm{Can}(x,\alpha) \supset (\mathrm{Does}(x,\alpha) \mathbin{\Box\!\!\rightarrow} q)]$	$\wedge$-intro, 10, 13
15	Mq	def, 14

4. Conclusions

Much of the discussion of the consequence argument has focused on beta-type principles. The present formulations allow those issues to be put to rest if we define the N operator appropriately. The big question in regard to the consequence argument does not concern beta-type principles, since beta-2 in an appropriate formulation is a theorem, but whether $N(P \wedge L)$ is true, i.e., whether we ever had, have or will have a choice about $P \wedge L$. It seems obvious to many that $N(P \wedge L)$ is true, but Lewis (1981) has argued against NL, and at least implicitly against $N(P \wedge L)$, and it may be more fruitful for the debate to return to the question of the truth of $N(P \wedge L)$ (see also Beebee 2002 for a similar conclusion), or of $M(P \wedge L)$ in the alternative version of the argument.[2]

5. Bibliography

Beebee, H. 2002. "Reply to Huemer on the Consequence Argument." *Philosophical Review* 111 (2002): 235–241.
Finch, A. and Warfield, T. A. 1998. "The MIND Argument and Libertarianism." *Mind* 127 (1998): 515–528.
Huemer, M. 2000. "Van Inwagen's Consequence Argument." *Philosophical Review* 109 (2000): 525–544.
McKay, T., and Johnson, D. 1996. "A Reconsideration of an Argument Against Compatibilism." *Philosophical Topics* 24 (1996): 113–122.
Pruss, A.R. 2007. "Conjunctions, Disjunctions and Lewisian Semantics for Counterfactuals." *Synthese* 156 (2007): 33–52.
Pruss, A.R. 2011. *Actuality, Possibility and Worlds.* London: Continuum.
van Inwagen, P. 1983. *An Essay on Free Will.* Oxford: Oxford University Press.
van Inwagen, P. 2000. "Free Will Remains a Mystery." *Philosophical Perspectives* 14 (2000), *Action and Freedom*: 1–19.
Widerker, David. 1987. "On an Argument for Incompatibilism." *Analysis* 47 (1987): 37–41.

[2]I would like to thank William Demsar for a number of discussions of the issues in this paper, as well as two anonymous readers and the editor, all of whose comments have significantly improved the paper.