Danilo Šuster
University of Maribor

Lehrer and the Consequence Argument*
Preprint

The consequence argument of van Inwagen is widely regarded as the best argument for incompatibilism. Lewis's response is praised by van Inwagen as the best compatibilist's strategy but Lewis himself acknowledges that his strategy resembles that of Lehrer. A comparison will show that one can speak about Lehrer-Lewis strategy, although I think that Lewis's variation is dialectically slightly stronger. The paper provides a response to some standard objections of incompatibilists to the Lehrer-Lewis reply.

Keywords: consequence argument, Lehrer, Lewis, compatibilism, power conditionals, explanatory conditionals

1. Introduction

In his work on free will, Keith Lehrer continues the grand tradition of classical compatibilism. I can not do full justice to his complex and fine theory (with loops and meta-levels) which has evolved over time. Basically, our preferences are the source of our freedom and since, according to Lehrer's analysis, the conditions for freedom of preference are compatible with determinism, freedom of action of the sort pertinent to moral evaluation is compatible with determinism. In this paper I shall discuss a classical *riposte* and Lehrer's reply:

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

This is an informal version of the consequence argument (CA), widely regarded as the best argument for incompatibilism – the conclusion that if determinism is true, then no one ever *really* has a choice about anything. Given van Inwagen's late comments one might think that Lehrer is more connected with the very introduction of the argument than with its rebuttal:

1

When van Inwagen got around to writing down the argument that had occurred to him in his conversation with Brand – he first did this in a doctoral thesis he wrote under the supervision of Richard Taylor (*de jure*) and Keith Lehrer (*de facto*) ... . Eventually, however, van Inwagen was able to publish two papers in which he argued for the incompatibility of free will and determinism. (He has always suspected that "The Incompatibility of Free Will and Determinism," which had been rejected by many journals, was accepted by *Philosophical Studies* only because Sellars chose Keith Lehrer as its referee, …) (van Inwagen 2004a, 215).

Well, with adversaries like that, who needs advocates? Still, van Inwagen (2008, 330) has plenty of praise for Lewis (1981), which is supposed to be: "… the finest essay that has ever been written in defense of compatibilism – possibly the finest essay that has ever been written about any aspect of the free will problem." Lewis's essay is a reply to van Inwagen's version of CA and in a footnote of this paper (Lewis 1986, 294, fn 3) we read: "Up to a point, my strategy here resembles that of Keith Lehrer."

Two questions emerge immediately. If Lehrer has already shown us a proper way to reply to the consequence argument, why is his (1980) not "the finest essay that has ever been written in defense of compatibilism"? A comparison will show that one can speak about a Lehrer-Lewis strategy, although I think that Lewis's variation is dialectically slightly stronger. And, secondly, how plausible is a Lehrer-Lewis reply to CA given some standard incompatibilistic objections? To put it briefly – it survives.

2. Lehrer and Lewis

Let me start with the following rough characterization: to say of someone that she "has free will" or "has a choice" with respect to a certain action at a given time is to grant her an ability to act in that way and an ability to do otherwise (van Inwagen 1983, 162). The compatibilists reply to CA that not all ways of being determined not to do something on a given occasion are ways which amount to *inability* to do it on that occasion. Expressions such as "it is up to the agent …," "to have a choice about …," do not have a sufficiently precise meaning. Suppose we explicate these notions in terms of a *preference* conditional: If a person would have acted in a certain way if she had preferred to do so, then that person could have

acted in the specified way (was free to act in that way). The premises of CA are true in this reading; laws (past) are not up to us because even if we had preferred the laws of nature (past) not to be as they were, they would still have been as they were. But the conclusion is false, an action of mine might well be determined, but still up to me. Had I preferred to act otherwise (and other conditions had been satisfied), I would have done so.

Non sequitur is a traditional compatibilist diagnosis of CA but conditional analysis of our ability to act otherwise is discredited both by van Inwagen ("no argument has ever been given in defence of the thesis that statements ascribing to an agent the power to act can be correctly analysed as conditionals" says van Inwagen 1983, 122) *and* by Lewis, as we shall see. Lehrer also rejects traditional compatibilists attempts to analyse 'could have' statements as conditionals, but he still offers a sophisticated preference conditional as a *sufficient* condition for the ability to act otherwise: If a person would have acted in a certain way if she had preferred to do so and certain other conditions about her preferences had been satisfied, then that person could have done otherwise (Lehrer 1980, 192). Here is his version of CA:

> Suppose that determinism is true and that S does not do *A* at *t.* Then at any time, *t',* before the birth of S, there was some condition, perhaps the state of the universe at *t',* such that it is a consequence of that state and the laws of nature that S does not do *A* at *t.* Therefore, if *S* could have done A at *t,* then *S* could have brought it about that either the laws of nature are different or the state of the universe is different at *t'.* But obviously *S* could not have brought about either of these conditions. So, S could not have done *A* at *t* (Lehrer 1980, 199).

The critical move introduced by Lehrer is to accept that there would *be* a difference but deny that the agent could have *brought* about the difference:

> The defect in the previous argument rests on the assumption that if *S* could have done *A* at *t,* then *S* could have brought about the conditions cited. If S had done *A* at t, then, of course, either the laws of nature would have been different or the state of the universe would have been different. But that is not to say that the person could have brought about these conditions (Lehrer 1980, 199).

The agent's inability to bring about the difference is then explained in terms of a false conditional (Lehrer 1980, 199): "… it is false that if S had preferred that either the laws of

nature not be as they were or that the state of the universe at *t* not be as it was, then one of these conditions would have been satisfied." Lewis (1986, 294) applauds the distinction:

> Lehrer grants a *weak* thesis: the agent could have done something such that, if he had done it, there would have been a difference in either laws or history. He rejects, as I would, the step from that to a *stronger* thesis: the agent could have brought about a difference in laws or history.

But he criticizes Lehrer's *explanation* of the agent's inability to bring about a difference:

> (according to Lehrer) … it is false that if the agent had preferred that there be a difference in laws or history, there would have been a difference in laws or history. I say, first, that this conditional may not be false. Suppose the agent is predetermined to prefer that there be no difference; had he preferred otherwise, there would have been a difference. (Had anything been otherwise than it was predetermined to be, there would have been a difference in either laws or history.) And second, if this conditional is not false, that is not enough to make the stronger thesis true. There must be some other reason, different from the one Lehrer gives, why the stronger thesis is false (Lewis 1986, 294).

One has to agree. Consider *The Age of Innocence* by Edith Wharton. In his past, Newland, the main character, has made a difficult but, in the long run, prosperous decision (not to leave his wife). Now, as an elderly man, he does not want that there be any differences in his past, but had he now preferred that there be a difference (to abandon his wife for Countess Ellen Olenska as he had once almost decided to do), then, given determinism, *that could only have been because* there would have been a certain difference in his past life and/or laws of nature resulting in his different state of present preference. In Lehrer's terminology – if Newland's preferences about the laws of nature or the state of the universe in his past had been different, then the necessary condition for this difference would (have to) have occurred. But it does not follow that Newland could have *brought about* this condition.

There is another problem with Lehrer's brief explanation of why it is false to say that by doing otherwise in a deterministic universe one could have brought about differences in the laws of nature or in the past. How should we understand the conditional starting with "if S

had preferred that either the laws of nature not be as they were or that the state of the universe at *t* not be as it was, ... ?" Is this a preference conditional which, according to Lehrer, states a *sufficient* condition for the ability to act otherwise? Well, sufficiency is not necessity, so the failure of this conditional would not, *by itself*, show that the agent lacked the ability to bring about the relevant difference.[1] Perhaps Lehrer, a prominent critic of traditional conditional analyses of ability, was still under the influence of a classical compatibilist's reply to CA?

Lewis offers an explanation in terms of a distinction between *weak* ability: "I have the ability to do something such that if I did it, a law would be broken" and *strong* ability: "I have the ability to do something such that if I did it, my act itself (or a consequence of my act) would cause (or be) a law breaking event." A free predetermined agent could have rendered a true proposition *L* specifying the laws of nature that govern our world false in the weak sense only. By acting otherwise as she actually did she is able to do something such that, if she did it, laws would have been falsified (though not necessarily by her act, or by any event caused by her act). The agent lacks strong ability with respect to the laws of nature.

Still, I think that the pair of distinctions corresponding to weak / strong: *L* would be falsified / *L* would be falsified by the agent's act itself (Lewis) or *L* would be falsified / a falsification of *L* would be brought about by the agent (Lehrer), is so similar that we can speak about *the Lehrer-Lewis* response to CA.

Incompatibilists reply to the Lehrer-Lewis analysis with an observation that *strong* must be the intended reading of ability in CA. Thus Kane (1996, 50):

> … it is clear to us that the proper reading of the premises of the Consequence Argument is the strong sense of 'can render false.' The only thing that could recommend the weak sense to anyone, as we see it, is a prior commitment to compatibilism.

Plausible enough. A man on the street, ignorant of philosophical technicalities, will understand his ordinary ability to raise his actually unraised arm in the causal reading as the ability to bring about a certain state of affairs. But now the compatibilist must face an immediate problem.

Suppose determinism is true, and Newland did not abandon his wife (¬*A*), but he could have done so (*A*). Let *H* stand for the proposition that expresses the state of the world at

---

[1] Thanks to the anonymous referee for pointing this out.

t_0, the distant past and *L* stand for the conjunction of the laws of nature. *Ability* may be expressed through a sentence operator such as '$A_{S,t}\,p$' abbreviating 'at time t, agent S is able to bring about that *p*' or simply '$A_S\,p$'. Van Inwagen uses the phrase "an agent can(not) render false a proposition that, ….", but he translates "He could have reached Chicago by midnight," as "He could have rendered the proposition that he did not reach Chicago by midnight false" (van Inwagen 1975, 189). So it looks unproblematic to understand "S could have rendered false a proposition that *p*" as '$A_S\neg p$'. Consider now the following simple "Diodoran" argument:[2]

1. $(H \,\&\, L) \rightarrow \neg A$ Assumption, determinism ('$\rightarrow$' stands for logical implication)
2. $A_N\, A$ Assumption, Newland could have acted otherwise
3. $A_N \neg(H \,\&\, L)$ 1, 2   *Master*

Master is the "Transference of Power" principle (Kane 1996, 47): "$A_S\neg q$, q is entailed by p, so $A_S\neg p$." Van Inwagen (1975, 192) uses the principle: "If S can render R false, and if Q entails R, then S can render Q false," which, he says, "seems to be *analytic*." The conclusion of this three-liner is that a free determined agent has a strong ability to bring about the falsity of the *conjunction* of history and law. Not easy to believe, so let us call this result STRANGE.

For incompatibilists STRANGE is incredible, and *Master* beyond doubt, so soft determinism, the doctrine that one is sometimes able to act otherwise as predetermined by past history and the laws of nature, implied by the joint truth of (1) and (2), must be false. Given the absurdity of STRANGE and the indubitability of *Master* the joint truth of (1) and (2) is not even possible, so compatibilism must be false.[3] Most compatibilists agree that STRANGE is incredible, but they blame *Master* (Perry 2004, Kapitan 2002). Thus Lehrer (2004, 64):

> There is a principle that figures into the argument in one form or another which says that if a person can do something, and something is a necessary condition for his doing it (like the falsity of either the past or the laws of nature), then, if the person can do the thing,

---

[2] I borrow the name from Kapitan (2002) who discusses a variety of logical principles for practical modalities, based on "Whatever is a consequence of a possibility is itself possible," ascribed to Diodorus Chronus.
[3] Thanks to the anonymous referee for this clarification.

then the person can bring about the necessary conditions for doing it as well. This is plausible, …, but false.

Lewis is an exception in the camp. He says explicitly that he cannot question the first four premises of van Inwagen's 1975 version of the argument and *Master* is one of them. Given that he (provisionally at least) accepts (1) and (2) this commits him to STRANGE, which he readily accepts (1986, 297, italics are mine):

> If I could have raised my hand despite the fact that determinism is true and I did not raise it, then indeed *it is true both in the weak sense and in the strong sense that I could have rendered false the conjunction HL of history and law*. But I could have rendered false the law proposition L in the weak sense, though I could not have rendered L false in the strong sense.

Lehrer would say that (2) in our simple Diodoran argument is true in the strong (causal) sense, but then a shift in the meaning of 'A_N' occurs, and (3) is true in the weak sense only. If Newland had abandoned his wife then the conjunction of the laws of nature and the state of the universe would have been different. But that is not to say that Newland could have brought about these differences. According to Lewis (3) *is* true in the strong sense, but no miracles (law-breaking events caused by Newland's actions) follow since he denies the following principle for strong ability: "$A_N \neg(H \& L) \rightarrow (A_N \neg H \vee A_N \neg L)$" (a version of agglomeration).[4] Counter-examples are discussed in the literature and easy to find. I am able to pick a red or black card from a pack of cards – by picking a card randomly I am strongly able to falsify a conjunction of not having a red card and not having a black card. It does not follow that I can render it false that I do not have a red card or that I can render it false that I do not have a black card.

Which way to go? A Lewisian position with its commitment to STRANGE looks unattractive, but I think there are *dialectical* advantages in accepting the *Master*. A compatibilist can simply bypass a typical table-tennis of examples, counter-examples and qualifications of the principle. She can grant that *Master* (with qualifications) looks at least *plausible*. So what? A monarchist is committed to defending a monarchy and a compatibilist is committed to (what seem to be) the analytic consequences of her thesis! Being able to act

---

[4] Define unavoidability of *p* for agent S, '$U_s p$,' as '$\neg A_s \neg p$'. The principle above then says that unavoidability is agglomerative: "$(U_N H \& U_N L) \rightarrow U_N (H \& L)$."

otherwise if determinism is true amounts to being able to falsify the conjunction of history and the laws of nature from which the actual action follows. If you do not believe it, you do not believe it (it is STRANGE, after all), but this comes very close to disbelieving compatibilism from the very beginning, and what then is the purpose of the consequence *argument*? Can you really persuasively *argue* that compatibilism is incredible because STRANGE, (a seemingly) analytic consequence of compatibilism, is incredible? Is this not just question-begging, like arguing: this man cannot be a bachelor since he cannot be an unmarried adult male?

3. Ability: two perspectives

A problem for the Lehrer-Lewis reply to CA remains, however. Wherever we place the weak / strong distinction, why does *weak* appear at all? Is this not just *ad hoc*, to defend compatibilism? Moreover, not only is our ability to *change* the laws of nature problematic, it looks bad enough for compatibilism if we had the power so to act that the laws of nature would have been different from the way they actually are. Even weak ability seems to be an ability to perform miracles if we define the ability to perform a miracle as the ability to bring about an event or state of affairs whose occurrence would be inconsistent with the whole truth about the past and the laws of nature (that is, basically, van Inwagen's reply to the Lehrer-Lewis solution, van Inwagen 2004b, 349).

Lehrer (2004, 67) addresses the second problem:

… it may be true of some but not all laws of nature, namely those concerned with matters other than human choice and preference, that they cannot be falsified by human choice or preference, but it does not follow from this that no law about human choice or preference can be falsified by human choice or preference.

An incompatibilist will reply that a law is a law, whether it figures in a causal explanation of human action or in a causal explanation of the movement of stars, so free will in a deterministic world implies the ability to perform miracles. I think that even a compatibilist will have to concede that there is something *paradoxical* in the ability of a free predetermined agent to act otherwise – to "falsify" laws about human choice and preference. In the same way, there is something paradoxical about our claims of ordinary knowledge

about the external world when faced with the challenge of radical skepticism. In solving the paradox it is not enough just to select some consistent stance on the propositions involved in a skeptic's argument. Proper treatment of the paradox will require one to explain also why it is that the skeptic's premises are so plausible. Consider:

P1   I did not raise my hand but I had the ordinary ability to do so. If so, and determinism is true, I have the ability to bring about an event or state of affairs whose occurrence would be inconsistent with the whole truth about the past and the laws of nature. Is this not the ability to perform a *miracle*?

P2   I did not raise my hand but I had the ordinary ability to do so. If so, and determinism is true, then the antecedent conditions (laws of nature, distant past) which implied that I would not raise my hand, would *have* to have been different. So what?

Let us take P1 and P2 as two *perspectives,* or points of view. The first perspective (P1) points to the implications of our actions from the active, practical, agent's point of view. The second perspective (P2) might be called the theoretical or explanatory point of view. Had we acted differently, some of the factors (which causally explained our actual action) would have been different. Is it not even more "crazy" to claim that had we acted differently, *nothing* would have been different? What would *explain* the difference?

We seem to have a case of "bistable" perception – same data ("I could have acted otherwise even if determinism is true"), but different perception ("miracle / so what?"). When a person stares at the picture of a *Necker* cube, it will often seem to flip back and forth between the two valid interpretations. The image leads to an alternation between two mutually exclusive perceptual states. Analogously, when we take the skeptical stance (suppose we are deceived by an evil demon), all knowledge disappears but when we take the Moorean stance, the ordinary knowledge claims look unassailable. And analogously, when one "stares" at the abilities of a free predetermined agent, one will flip back and forth between two perspectives based on two interpretations of the crucial conditional.

If determinism is true and Newland did not abandon his wife, then at least one of these counterfactuals is true:

If Newland had abandoned his wife, the remote past would have been different (would *have* to have been different).

9

If Newland had abandoned his wife, the laws would have been different (would *have* to have been different).

Some compatibilists prefer the first conditional (Perry 2004); Lewis defends the second. Lehrer's earlier writings suggest that a compatibilist should accept the fixity of the laws of nature (Lehrer 1990). But in his latest work (Lehrer 2004) he is open to the idea that laws concerning our choices and preferences would have been different had we acted otherwise. The incompatibilist understands both of these counterfactuals "actively," as *power* conditionals on par with: If I had pushed the switch, the light would have gone off. Assuming that determinism is true, would my action not in a similar way miraculously make a difference in the laws of nature (past)?

The "offending" counterfactuals are backtrackers, conditionals containing antecedents concerned with events or states of affairs occurring or obtaining at times later than those involved in the consequents of the conditionals (a difference in laws is explained by a *prior* "divergence miracle" according to Lewis). It is widely held that such conditionals are rarely true, but sometimes they are. Note the difference in phrasing indicated in the brackets (cf. also Peacocke 1999, 326):

If A had been the case, then B would have been the case.

If A had been the case, then B would *have* to have been the case. Or: If A had been the case, then that *could only be because* B had been the case at the earlier time.

In the "only because" conditional, the consequent (usually) offers an *explanation* for the antecedent of the conditional, so I will speak about explanatory conditionals. Thus Bennett (2003, 273):

"If the plane had arrived at 2:00 p.m., it would have left at noon."

is more naturally and happily worded like this:

"If the plane had arrived at 2:00 p.m., it would have to have left at noon."

The consequent is acceptable as the best *explanation* for the antecedent. If the plane arrived at 2 p.m., that would have to have been because it left at noon—the modal "have to" expresses the compulsion in our being forced to that explanation for lack of any other as good.

This pattern does not imply any miraculous powers on the part of a free determined agent. The crucial conditional is an *explanatory* conditional, not a *power* conditional. The introduction of *weak* marks this distinction. Suppose determinism is true, I did not raise my hand, but was free to do so. A Lewisian compatibilist will say:

If I could have raised my hand (and thereby brought about the falsification of the conjunction of history and laws), then I could have rendered the conjunction of laws of nature false. That is to say: I could have done something which could only be explained by an appropriate difference in the laws of nature.

And a compatibilist who denies *Master* will say:

If I could have raised my hand, then I could have rendered the conjunction of history and laws false. That is to say: I could have done something, which could only be explained by an appropriate difference in the conjunction of laws and history.

We have a combination of two perspectives – *weak* ability (theoretical, P2) in the consequent functions as an explanation of *strong* ability to raise one's hand (active, P1) in the antecedent. The incompatibilists will of course protest, thus Ekstrom (2000, 49): "The only factor recommending a switch in reading to the weak sense of ability to render false for the consequent of the conditional … is, it seems, a prior commitment to compatibilism." We can understand this objection as a requirement of uniform perspective – *active* should figure both in the antecedent and in the consequent.

But the crucial transfer principle for strong ability (*Master* if you go the Lehrer way or agglomeration if you go the Lewis way) is invalid, so the switch *is* motivated, after all. No incredible powers are *transferred* from the antecedent (ability to raise one's hand) to the consequent (ability to falsify laws/history). The incompatibilist will still protest – even if *weak* admittedly makes sense, it should nevertheless be understood actively, as a sort of ability to *make* the difference and not just indicating something that *explains* the difference. A

requirement of uniform perspective should be respected. According to the Lehrer-Lewis proposal, there is an active, first person point of view of a free agent in the antecedent and a theoretical, third person point of view of an agent as the object of causal and motivational histories in the consequent of the conditional. Is this combination of perspectives not paradoxical, like a familiar impossible object based on an inconsistent interpretation of the Necker cube? [5]
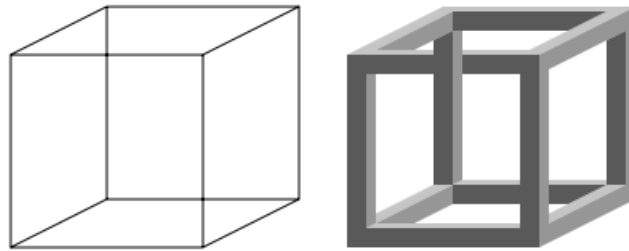
But this combination, a compatibilist will have to say, that ties together our conception of ourselves as free with our conception of a causal order explaining our freedom (Lehrer 2004, 68) is precisely our situation, the human condition. There is an air of paradox in the alternation between two mutually exclusive perspectives but the alleged impossibility is just an illusion.

---

[5] A picture from *Wikipedia*: http://en.wikipedia.org/wiki/Impossible_cube

* Presented on the Workshop "The Philosophy of Keith Lehrer", Graz, 21. - 23. October 2010.

References

Bennett, J. (2003). *A philosophical guide to conditionals*. Oxford: Oxford University Press.

Ekstrom, L. (2000). *Free will: a philosophical study*. Boulder, Col.: Westview Press.

Kane, R. (1996). *The significance of free will*. New York: Oxford University Press.

Kapitan, T. (2002). A Master Argument for Incompatibilism? In Kane, R. (Ed), *The Oxford Handbook of Free Will* (pp. 127-157). Oxford: Oxford University Press.

Lehrer, K. (1980). Preferences, Conditionals and Freedom. In van Inwagen, P. (Ed), *Time and cause: essays presented to Richard Taylor* (pp. 187-201). Dordrecht: D. Reidel.

Lehrer, K. (1990). A Possible Worlds Analysis of Freedom. In Lehrer, K. *Metamind* (pp. 43-78). Oxford: Clarendon Press.

Lehrer, K. (2004). Freedom and the Power of Preference. In Campbell, O'Rourke, Shier, (Eds.), *Freedom and Determinism* (pp. 47-69). Cambridge MA: MIT Press.

Lewis, D. (1981). Are We Free to Break the Laws? *Theoria* 47(3), 113-21. Reprinted in Lewis, D. (1986). *Philosophical Papers II* (pp. 291–298). New York and Oxford: Oxford University Press.

Peacocke, C. (1999). *Being Known*. Oxford: Oxford University Press.

Perry, J. (2004). Compatibilist options. In Campbell, O'Rourke, Shier, (Eds.), *Freedom and Determinism* (pp. 231–254). Cambridge MA: MIT Press.

van Inwagen, P. (1975). The Incompatibility of Free Will and Determinism. *Philosophical Studies* 27(3), 185-99.

van Inwagen, P. (1983). *An essay on free will*. Oxford: Clarendon Press.

van Inwagen, P. (2004a). van Inwagen on free will. In Campbell, O'Rourke, Shier, (Eds.), *Freedom and Determinism* (pp. 213-230). Cambridge MA: MIT Press.

van Inwagen, P. (2004b). Freedom to break the laws. *Midwest Studies in Philosophy* 28(1), 334-350.

van Inwagen, P. (2008). How to think about the problem of free will. *Journal of Ethics* 12, 327-341.