Croatian Journal of Philosophy
Vol. XXI, No. 63, 2021
https://doi.org/10.52685/cjp.21.63.2
Received: June 6, 2021
Accepted: November 15, 2021

Arguing about Free Will: Lewis and the Consequence Argument

DANILO ŠUSTER
University of Maribor, Maribor, Slovenia

I explore some issues in the logics and dialectics of practical modalities connected with the Consequence Argument (CA) considered as the best argument for the incompatibility of free will and determinism. According to Lewis (1981) in one of the possible senses of (in)ability, the argument is not valid; however, understood in the other of its possible senses, the argument is not sound. This verdict is based on the assessment of the modal version of the argument, where the crucial notion is power necessity ("no choice" operator), while Lewis analyses the version where the central notion is the locution "cannot render false."Lewis accepts closure of the relevant (in)ability operator under entailment but not closure under implication. His strategy has a seemingly strange corollary: a free predetermined agent is able (in a strong, causal sense) to falsity the conjunction of history and law. I compare a Moorean position with respect to radical skepticism and knowledge closure with ability closure and propose to explain Lewis's strategy in the framework of his Moorean stance.

Keywords: The consequence argument; compatibilism; (in)ability; closure; radical scepticism; Moorean stance.

> Philosophy is certainly *not* self-expression. … philosophy, of course, is argument, and you can say, well, is the conclusion true and is the argument valid? Iris Murdoch[1]
>
> How can philosophical enquiry be conducted without a perpetual *petitio principii*? Frank Ramsey[2]

[1] Quoted from Setiya (2020: 66).
[2] Ramsey (1994: 2).

1. *Introduction*

When arguing about free will, the consequence argument (CA), also called "the" argument for incompatibilism, a "master" argument, or the "unavoidability argument", is widely regarded as the best argument for the incompatibility of free will and determinism. Briefly, the remote past and the laws of nature are not up to us. However, if determinism is true, each of our present actions is a consequence of the laws of nature and the remote past. Therefore, the consequences (including our present acts) are not up to us either, and no one enjoys free will (cf. van Inwagen, 1983: v).

The evaluation of this piece of reasoning as an *argument* depends on two questions. What does it mean that one "thing" is a *consequence* of another? Do we rely on implication or something stronger, logical consequence (entailment)? And how to evaluate the *premises*: what does it mean to say that something is (not) up to us? Sheer inescapability, lack of causal control, or something different altogether? What are the starting points in this debate? What is the role of *logic* in establishing the final conclusions about CA?

I will approach these questions by focusing on Lewis's (1981) reply to van Inwagen's version of the argument. According to Speak (2011: 115), "the single most influential contribution to the overall philosophical quality of the recent free-will debate is van Inwagen's careful development of what he has dubbed the 'Consequence Argument.'" Van Inwagen, on the other hand, compliments Lewis (1981) as "…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" (van Inwagen 2008a: 330). Well, Lewis's article is a *refutation* of CA, so, at the end of the day, whom should we praise—defense or prosecution?

The logical core of CA is rule *Beta*, the transfer of powerlessness: roughly, "It is unavoidable that p and it is unavoidable that (if p, then q); hence, it is unavoidable that q." Since the past and the laws of nature are (supposed to be) fixed and unavoidable, it follows, via determinism (the past and the laws of nature together determine everything) and a Beta-like *transfer* principle, that each of our choices is unavoidable. Principle Beta officially entered the philosophical scene with the third (modal) version of CA (van Inwagen 1983: 93–105). However, Lewis (1981) discusses the original, non-modal version of CA (van Inwagen 1975), which is silent on Beta. So, how does Lewis criticize CA? Does he object to its inferential structure, or does he deny the premises? Is he a Beta "blocker" or does he use some other "pain killer" to block CA?

Although the topic has been much discussed, these questions have not been clarified in a satisfactory way. The answer is not simple. Lewis introduces two senses of ability, and a parallel distinction between the two senses of unavoidability nowadays constitutes the core of what is called the *main compatibilistic response* to CA. Understood in one of

the possible senses ("weak"), the laws of nature are not unavoidable, which makes Lewis a "fixity finesser" (Speak 2011: 121). Understood in the other of its possible senses ("strong"), rule Beta is invalid: the premises of CA are true, yet we are still able to act otherwise. In order to properly assess the question of (in)validity, we must enter into a discussion of modal principles and practical modalities—a terrain well-traversed (cf. Kapitan 1991, 1996, 2002 and 2011). Nevertheless, I hope to add new details and emphasize points that have not so far been noted. Lewis accepts closure of the relevant (in)ability operator under *entailment* but not closure under *implication* (some of the results of a slightly more technical nature are given in the *Appendix*).

Closure under entailment has a strange corollary. It is usually taken for granted that strong ability is *causal*. Suppose I have just put my hand down on my desk and this was a free but predetermined act. I could make it the case that my hand was raised. According to Lewis, if determinism is true, then the very act of raising my hand or some event caused by this act would *directly* falsify any sufficiently inclusive conjunction of history and law. To quote Lewis on a slightly different issue (1986: 292): "A marvelous power indeed! And with so little effort!" The incompatibilists view this result as a refutation of Lewis's response to CA and compatibilism in general, while some compatibilists tend to either neglect or deny it explicitly (Rummens 2019, Perry 2004, Kapitan 2011). Why would Lewis accept such a consequence? In this paper, I seek to *defend* Lewis's original strategy. It is not easy to resolve all the tensions generated by Lewis's interpretation but some recently published material sheds new light on this debate (Lewis et al. 2020 and Lewis 2020). From the general perspective of a Moorean stance on the free will issue (it's a Moorean fact that we often have a choice what to do), the strange corollary is plausible and defensible. I end the paper with some comments on the role of *logic* and *dialectics* in establishing the conclusions about CA.

2. *CA—the modal version*

Let me start with some standard terminology and notation: '□' stands for *broad* logical necessity (including metaphysical necessity); '→' is a sign of the material conditional; 'H' stands for a true proposition about the total state of the world at some moment in the distant past[3]; 'L' designates the conjunction of the laws of nature, and 'p' an arbitrary true proposition about the present or future (typically about an action, "S raised her hand at *t*"). A special modal operator 'N' is introduced to express "power" necessity, as opposed to free will. Van Inwagen (1983: 93) defines 'Np' as: *p* is true and no one has or ever had any choice about *p*. Determinism ('DET') is the thesis that the past (a complete specification of the universe at any given instant in the past) and the

[3] Van Inwagen uses 'P_0', but to be consistent with Lewis, I will use 'H' throughout.

laws of nature together determine everything. Two principles of inference are used:

Alpha $\Box p \vdash Np$
Beta $Np, N(p \rightarrow q) \vdash Nq$

The modal argument or the third version of the consequence argument (CA_3) is then:

(1)	$\Box [(H \& L) \rightarrow p]$	// DET (premise)
(2)	$\Box [H \rightarrow (L \rightarrow p)]$	// PL, 1
(3)	$N [H \rightarrow (L \rightarrow p)]$	// Alpha, 2
(4)	N H	// No choice about the past (premise)
(5)	$N (L \rightarrow p)$	// Beta – (3), (4)
(6)	N L	// No choice about the laws (premise)
(7)	N p	// Beta – (5), (6)

If determinism is true, then it is not within anyone's power to perform any actions other than those they do perform. Van Inwagen admits that Beta is the most challenging element of the argument to defend and the "most doubtful thesis the incompatibilist must accept" (van Inwagen 1983: 222). This proved to be the case: there are counter-examples to the original Beta, acknowledged by van Inwagen, who later wrote (2017: 118):

> I mistakenly supposed that the only way in which it could be that one had no choice about the truth-value of a proposition would be for the truth-value of that proposition to be in some way so firmly "fixed" that one was unable to change it. I did not see that there is another way for one to have no choice about the truth-value of a proposition: for that truth-value to be a mere matter of chance.

If we understand having a choice about the truth value of a proposition as being able to reliably *ensure* a certain result, we face the problem of chancy events not under our control. Suppose I do *not* toss a coin but could have done so. Let *p* be "the coin does not land heads" and *q* stand for "the coin does not land tails". It is then true that 'Np' (nobody can act to ensure that the coin lands heads); it is also true that 'Nq' (nobody is able to act to *ensure* that the coin lands tails). However, it does not follow that 'N (p & q).' This conjunction, "the coin does not land heads and the coin does not land tails" is true, but by flipping a fair coin, I could ensure that the coin lands tails or heads, so I am able to render the conjunction 'p & q' false. This is a counter-example to the principle of Agglomeration (McKay and Johnson 1996: 115):

$$Np \& Nq \vdash N (p \& q)$$

With Alpha impeccable, the invalidity of Agglomeration implies the invalidity of Beta (for details, see *Appendix*). To take our previous example: let *r* stand for "the coin is not tossed." It is still true that 'Np', but so is 'N (p → r)'. Nobody can ensure that "(p & ~r)" is the case (the coin does not land heads when tossed). Yet 'Nr' is false: I was able to toss the coin (cf. Carlson 2000: 283–84).

Faced with this type of counter-example to Beta, what logical options do incompatibilists have to amend CA? One could vary the interpretation of 'N', the type of consequence relation, or both. The truth value of a proposition can be beyond one's control because it is a matter of chance ("I am not able to win a fair lottery") or because it is fixed no matter what anyone does ("I am unable to construct a perpetuum mobile"). In the first case, although there is nothing I can do that *would* ensure the result, I *might* still (by sheer luck) achieve the happy outcome. In the second case, the laws of nature prevent me from doing anything that *might* possibly result in such a machine. The incompatibilists now usually opt for the second, "no matter what" notion of unavoidability (as suggested by McKay and Johnson, 1996: 118–119). According to Huemer (2000: 538), 'Np' is to be understood as "no matter what you do (among the things that you are able to do), *p*." Van Inwagen (2008b) later interprets 'Np' as "untouchable": *p* is true and no human being is or ever has been able to act in such a way that, if he or she did act that way, *p* might be or might have been false. This interpretation blocks the McKay-Johnson counter-example ('Np' and 'Nq' are now false; the coin *might* land heads when tossed). However, laws of nature and the distant past (premises 4 and 6 of CA_3) are supposed to be untouchable, true no matter what, so CA_3 remains a sound argument.

Are there any other options to revise CA? We might use the original "no choice" as "not being able to ensure the falsity of" interpretation for 'N' and proceed *directly*:

(1) $\Box [(H \,\&\, L) \rightarrow p]$ // DET (premise)

(2) $N\,(H \,\&\, L)$ // No choice about the past and the laws (premise)

(3) $N\,p$ // Beta 2 – (1), (2)

A different (weaker) rule is now used, *Beta 2*[4], which can be expressed as:

$$Np, \Box (p \rightarrow q) \vdash Nq$$

or,

$$\Box (p \rightarrow q) \vdash Np \rightarrow Nq$$

Logicians would describe the original Beta as closure of the relevant operator ('N') under *implication* and Beta 2 as closure under *entailment* or *simple* consequence: if *q* is a consequence of *p*, then N*q* is a consequence of N*p*. A McKay-Johnson type of counter-example to Beta is inefficient against Beta 2. Rewrite the previous example in the new form: 'Np, $\Box$ (p → r), so Nr'. The second premise is now false: it is possible, in a broadly logical way, that the coin does not land heads when tossed. The counter-example to Beta fails.

How about CA? A weaker rule requires a stronger premise. And now there remains one premise only: the conjunction '(H & L)' offers

[4] As designated by Warfield and Finch (1998), introduced by Widerker (1987) as *β'* but also called *Beta box*, *alpha 2* (Huemer 2000) and *T2* (Carlson 2000).

a description of what Finch and Warfield (1998: 523) call the "broad past", the complete state of the world at a time in the distant past, including the laws of nature. They maintain that the broad past is fixed (nobody can act so as to *ensure* that it is false), so the second premise of this version of CA is true. With Beta 2 impeccable, the argument is sound and the thesis of incompatibilism established.

To summarize: the original version of CA is based on the closure of the "no choice" operator ('N') under *implication*, but closure fails for this operator. A defender of a revised CA can use a stronger interpretation of power necessity or closure of the relevant operator under *entailment*. We now turn to Lewis's treatment of CA. How does he assess these rules and CA in general?

3. *Lewis and CA*

Lewis discusses the first non-modal version of the consequence argument (CA_1), and it is not immediately obvious where to locate the relevant transfer principles. In this version, the crucial notion of (in)ability is captured by the phrase "an agent can(not) render false a proposition that, …," describing abilities in terms of the agent's power "over" the truth-values of propositions. We start with the scenario in which a certain judge J did *not* raise his hand at the moment T, but could have done so, given the usual understanding of our abilities. *H* (the proposition that expresses the total state of the world in the remote past) and *L* (the conjunction into a single proposition of all laws of physics) are used as before. Instead of 'p' (an arbitrary true proposition), we now have 'P' denoting the proposition that expresses the state of the world at T. The argument is then (van Inwagen 1975: 191):

1. If determinism is true, then the conjunction *H* and *L* entails *P*.
2. If J had raised his hand at T, then *P* would be false.
3. If (2) is true, then if J could have raised his hand at T, J could have rendered *P* false.
4. If J could have rendered *P* false, and if the conjunction of *H* and *L* entails *P*, then J could have rendered the conjunction of *H* and *L* false.
5. If J could have rendered the conjunction of *H* and *L* false, then J could have rendered *L* false.
6. J could not have rendered *L* false.
7. If determinism is true, J could not have raised his hand at T.

J's (not) raising his hand is an arbitrary action of an arbitrary agent. So, if determinism is true, then it is not within anyone's power to perform any actions other than those they do perform.

The most plausible candidate for stating a Beta-like transfer principle of inference in this version is the fourth premise. According to van Inwagen, this principle seems to be *analytic*:

> This premise may be defended as an instance of the following general principle: If S can render *R* false, and if *Q* entails *R*, then S can render *Q* false (van Inwagen 1975: 192).

In the third version, we had "having no choice about"; in the first version, the central notion is "can(not) render false a proposition." To compare them, it is convenient to use the general logical framework of *practical* modalities. According to van Inwagen (1975: 189), we can translate "He could have rendered the proposition that he did not reach Chicago by midnight false" as "He could have reached Chicago by midnight," understood as the ascription of *ability* to an agent. We might then understand, "S could have rendered false a proposition that *p*" as, broadly, "S was able to act so that *p* would or might be false" and express this locution in terms of a modal operator of ability indexed to agents as 'A_S~p'. The contemporary framework of agency-centered modalities is called the *stit* theory. The agent α *s*ees *t*o *i*t *t*hat *A* is defined as true at a certain moment *m* of the world history *h* (index), just in case the action performed by α at that index guarantees the truth of *A*. The action might result in a variety of possible outcomes, but the statement *A* must be true in each of them. I will follow Kapitan (2011: 131) and abbreviate the agent's ability to see to it that a situation *p* obtains as 'Ap,' their inability to see to it that *p* as '~Ap' and their inability to prevent *p* as '~A~p.'[5]

We can now spell out premise (4) of CA_1 in terms of the ability operator:

$$[A\sim P \,\&\, \Box\,((H \,\&\, L) \rightarrow P)] \rightarrow A\sim (H \,\&\, L)$$

This premise can be viewed as an instance of a more general logical principle for practical modalities (*Master*)[6]:

$$A\sim q, \Box\,(p \rightarrow q \vdash A\sim p$$

Recall that 'Np' says that *p* is true and that no one has had any choice about *p*. It seems natural to interpret having no choice as the inability to prevent *p*, so we get '*p* & ~A~*p*' as a plausible translation for 'Np'. There are difficulties with directly implementing this translation scheme (see *Appendix*), and different understandings of "ability" will resonate importantly in our further discussion. Still, this translation is consistent with Lewis's understanding of ability (but see discussion below). The most important result is a connection between the transfer principles for practical modalities and power necessity for our present purpose. Given the translation scheme and Lewis's understanding of ability, *Master* implies Beta 2. This is easy to see (details in the *Appendix*): If *p* entails *q*, and the ability to render *q* false implies the ability to render *p* false (*Master*), then the *in*ability to render *p* false implies

[5] Strictly speaking, we are talking about S's ability at time t, '$A_{s,t}$p', but skipping the indices will do no harm in my discussion.

[6] 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.

the *in*ability to render *q* false. It is then a short step to infer that the unavoidability (power necessity) of *p* implies the unavoidability of *q*. So, when *p* is unavoidable and *p* entails *q*, then *q* is also unavoidable, which is just Beta 2:

$$A\sim q, \Box(p \rightarrow q) \vdash A\sim p$$

implies[7] the validity of Beta 2 (closure of power necessity under entailment):

$$Np, \Box(p \rightarrow q) \vdash Nq$$

Let us return to Lewis and his assessment of CA_1. Lewis reformulates the argument for incompatibilism as a *reductio*: if we attribute ordinary abilities to agents in a deterministic universe (the judge J could have raised his hand), we are forced to credit them with a magical past or law-changing abilities, as well. This result is absurd, so, per *reductio*, if determinism is true, nobody has the ordinary ability to act otherwise. In his reply, Lewis launches a *distinguo* between two senses of ability (1986: 297):

> I could have rendered a proposition false in the *weak* sense iff I was able to do something such that, if I did it, the proposition would have been falsified (though not necessarily by my act, or by any event caused by my act).
> I could have rendered a proposition false in the *strong* sense iff I was able to do something such that, if I did it, the proposition would have been falsified either by my act itself or by some event caused by my act.

How does weak/strong distinction work as an antidote to CA_1? Here is a longer passage, emphasizing two points that are crucial for my reading:

> I did not raise my hand; suppose for *reductio* that I could have raised my hand, although determinism is true. Then it follows, ***given four premises that I cannot question***, that I could have rendered false the conjunction *HL* of a certain historical proposition *H* about the state of the world before my birth and a certain law proposition *L*. If so, then I could have rendered *L* false. (Premise 5.) But I could not have rendered *L* false. (Premise 6.) This refutes our supposition.
> To this, I reply that Premise 5 and Premise 6 are not both true. Which one is true depends on what van Inwagen means by "could have rendered false" (Lewis, 1986: 296). (…)
> 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. So, if we take the weak sense throughout the argument, then I deny Premise 6. If instead we take the strong sense, then I deny Premise 5 (Lewis 1986: 297).

Lewis treats premise (4) as uncontroversial, but this premise is an instance of *Master*, so we can ascribe to him the acceptance of Beta 2 (implied by *Master*). Nevertheless, *reductio* fails because there is no

[7] Kapitan (2011: 134) argues for a general claim: the closure of power necessity and ability closure are equivalent. I, however, prefer a more modest claim.

uniform reading of "could have rendered false" to make all the premises true. Surprisingly, premise (6) is false according to the weak reading of ability: had I raised my (actually unraised) hand, a law would have been broken beforehand and not by my act itself or by some event caused by my act. Even more surprisingly, the implication (5) is false according to the strong reading of ability. The act of raising my (unraised) hand would directly, *by itself*, falsify any sufficiently inclusive conjunction of history and law. So, the antecedent of (5) is true.

In contrast, the consequence of (5) is false: the act of raising my (unraised) hand would not falsify laws of nature in the strong sense. It would be preceded by "the divergence miracle" that would falsify a law (Lewis 1986: 297). The laws of nature, just prior to my acting, would be slightly different from the way they are. The alternative action would take place in a different possible world: it would represent a miracle *relative* to the laws of our world but remain lawful in the possible world where I act otherwise (cf. Rummens, 2019).

4. *The main compatibilist response*

The weak/strong distinction is usually interpreted as the core of *the main compatibilist response* (or 'MCR') to CA. The notion of ability is explicated in terms of a conditional, where there are two general readings, stronger and weaker. S is *broadly* able at t to see to it that p iff there is a course of action X such that at t (i) S is able at t to do X, and (ii) were S to do X, then p (Kapitan 2011: 135). No connection is specified between the course of action X and p. Another option is a narrower, causal reading: S is *causally* able at t to see to it that p iff there is a course of action X such that at t (i) S is able at t to do X, and (ii) S's doing X would make it the case that p. According to this *active* reading, the agent brings about or causes p to obtain; p is the case because of what the agent does; her action leads to or results in p. Causal ability implies broad ability but not vice versa (for instance, S is broadly able to see to it that a certain tautology is true, but the tautology is not true *because* of her actions).

The notion of weak or broad ability opens the space for the denial of fixity of laws (or the past): the agent is (broadly) able to act otherwise without having the *causal* ability to break a law of nature. A free predetermined agent could have done something such that, had they done it, there would have been a difference in either law or history. "But that is not to say that the person could have brought about these conditions" (Lehrer 1980: 199). Lewis's distinction weak/strong is then interpreted along the lines of broad/causal and applied to the third, *modal* version of CA. The critical move is to accept that the agent is broadly able to act in a way that would falsify the laws of nature or the past but deny them any stronger "causal" powers. The agent could not have *brought about* the relevant differences, so premise (6) in the third version of CA is false (and so is premise (6) in the first version). Alternatively, one

might question premise (4) of CA_3, if one makes a difference between an agent who has the ability to act in such a way that they *alter* the past, as opposed to an agent who has (broad) ability to act in such a way such that, if they did so act, the past would have been different.

With two notions of ability, we now have two options for the notion of power necessity in CA_3, so let us introduce '$N_w p$' (a dual of weak ability) for, roughly, "*p* is true and no act of any person could in *any* way (weakly or strongly) falsify *p*", and '$N_s p$' (a dual of strong ability) for "*p* is true and no act of any person could strongly falsify *p*." According to MCR, premise (6) of CA_3 (and CA_1) is false for '$N_w p$', but the transfer principle (Beta) fails if we adopt the causal (strong) reading '$N_s p$':

> ... if we adopt the broad sense of ability, then, although the argument is valid, at least one of its premises is false, whereas if we adopt the causal sense of ability, then, although the premises are correct, the argument is invalid because the relevant closure principle fails. As the most prominent exponent of this line of reasoning, David Lewis, concluded, there is no one consistent reading of the consequence argument for incompatibilism critical modality that would render the Consequence Argument sound. (Kapitan 2011: 138)

Van Inwagen (2008b: 455) agrees:

> The philosopher David Lewis has contended … that our technical term 'untouchable' is ambiguous, and that if the word is understood in one of its possible senses, the Conditional Rule [i.e. *Beta*] is invalid, and, if it is understood in the other of its possible senses, *L* is not an untouchable truth.

Why is Beta not valid? Premises (4) and (6) of CA_3 are true in the *strong* sense: the laws of nature and events in the distant past are "not up to us"; they are beyond our causal control. Not so for our actions, which are up to us in the sense of being brought about by our desires, abilities, character, and beliefs (cf. Slote 1982; Mele 2006: 138). So, the conclusion is false: as free predetermined agents, we are still *causally* able to act otherwise. However, premises (4) and (6) of CA_3 are false in the weak sense of ability. As free predetermined agents, we do not possess the causal ("strong") ability with respect to the past and the laws of nature; instead, either the past or the laws of nature would have been different (would have to have been different) for our free action to take place.

So, CA is a philosophical failure according to MCR: either not valid (Beta fails) or not sound. This may be the central line of MCR, but there are immediate problems with the projection of Lewis's weak/strong onto broad/causal. First, it is difficult to directly extract the "not valid or not sound" verdict on Beta and CA_3 from Lewis (1981) on CA_1, since this version is not based on Beta. Moreover, even *weak* is defined in terms of counterfactual sufficiency (cf. Huemer 2000: 529): S can render *p* false iff S can perform some action, such that were he to do so, it would not be the case that *p*. Had I acted otherwise, a certain result would be *reliably* correlated with my alternative action, directly (strong) or indirectly (weak). Nevertheless, we cannot "ensure" the out-

come of a chancy process, so the McKay—Johnson counter-example to Agglomeration works for both, for '$N_w p$' and '$N_s p$'.

Lewis's verdict on premises (4) and (5) of CA_1 reveals his assessment of the relevant transfer principles. Slote has noted that premise (5) "If J could have rendered the conjunction of *H* and *L* false, then J could have rendered *L* false" corresponds to the principle of Agglomeration for 'N'. Given the law of contraposition and the assumption of the fixity of the past, (5) is tantamount to "if one can't falsify a law of nature *L* and can't falsify a proposition about the past *H*, one can't render false the conjunction of *L* and *H*", which is just an Agglomeration assumption for the necessity expressed by "can't render false" (Slote, 1982: 10, fn 7). With Lewis, we now have two options for the relevant notions of necessity, so let me start with Agglomeration for the dual of strong ability:

$$N_s H \,\&\, N_s L \vdash N_s (H \,\&\, L)$$

Lewis clearly *denies* the validity of this pattern. The premises are true: "… if I had raised my hand, the intrinsic state of the world long ago would have been no different", neither do we possess *strong* ability with respect to the laws of nature. However, the conclusion is false—I could have rendered false the conjunction of history and law 'H & L' in *both* senses. A denial of Agglomeration implies a denial of Beta for 'N_s' (Slote indicated this already). Nevertheless, Lewis accepts premise (4) of CA_1 in both senses (the premise he cannot question), so he accepts Beta 2 according to the translations proposed. This is consistent, since Beta 2 and Agglomeration (plus Alpha) together entail the validity of Beta, so a denial of Agglomeration will block Beta (see Appendix). One can have closure of an operator under *entailment* without the corresponding closure under *implication*.

How about weak ability and the corresponding transfer principles? Replacing practical modalities with the appropriate notion of power necessity, contraposing (5) of CA_1 and using propositional logic, we get:

$$N_w L \rightarrow N_w (H \,\&\, L)$$

Can we add '$N_w H$' as a separate conjunct in the antecedent? Is the past completely isolated from our actions in both senses of ability? Not the entire past: under determinism, the counterfactual worlds in which I act otherwise must involve a difference (a divergence miracle) in the past immediately before my action. Still, I think that Lewis would agree with the *absolute* impossibility of rendering false a true proposition about the *remote* past, so we can insert this proposition as a second conjunct in the antecedent to get:

$$(N_w L \,\&\, N_w H) \rightarrow N_w (H \,\&\, L)$$

Agglomeration for 'N_w' looks acceptable in *this special* case at least. The conditional is at least *plausible*, given the content of the propositions involved. Fischer, in defense of his conditional version of the consequence argument, pointed out that sometimes the argument is not

formally valid, but it is nonetheless reasonable to accept its conclusion given the *content* of its premises (Fischer, 1994: 228, fn 43). Since Beta 2 (accepted by Lewis in both senses) and Agglomeration together entail the validity of Beta, we might accept Beta for 'N_w' as at least "materially" valid. Of course, in this case "N_wL" is false according to Lewis, and the modal argument (CA_3) based on the notion of 'N_w', a dual of weak ability, is not sound, precisely as diagnosed by the MCR.

5. *Problems*

Have we now confirmed the canonical interpretation of Lewis and established a correspondence between Lewis's weak and strong and the notions used by the main compatibilistic reply, broad and causal? Cracks begin to appear when we reconsider premise (4) of CA_1:

$$[A\neg P \,\&\, \Box ((H \,\&\, L) \rightarrow P)] \rightarrow A\neg(H \,\&\, L)$$

According to Lewis, this premise is valid whichever notion of (in)ability we consider. And he readily accepts the result (Lewis, 1986: 297): "My act of raising my hand would (by *itself*) falsify any sufficiently inclusive conjunction of history and law." This is not easy to accept, even less so when we interpret strong ability as *causal* ability, the ability to bring something about or make changes. This seems to be suggested by Lewis's examples of *strong* ability. For instance (Lewis 1986: 294):

> Therefore, I am able to break a window, a promise, or a law only if I am able to do something such that, if I did it, my act either would cause or would be a window-, promise-, or law-breaking event.

He then goes on to deny that a free predetermined agent possesses such an ability with respect to the laws of nature (the act of raising his hand would not itself be or cause a law-breaking event, rather a law would be broken beforehand if he did it). But, according to (4), a free predetermined agent possesses such an ability with respect to the *conjunction* of history and law. He could have rendered the conjunction false in the strong sense, so he was able to do something such that, if he did it, the conjunction would have been falsified either by his act itself or by some event caused by his act. His act itself would cause or be a conjunction-breaking event.

Let me use ODD for the claim that a free, predetermined agent possesses a *strong* ability to render false the *conjunction* of the remote past and the laws of nature. ODD, in general, does not look very attractive. The incompatibilists declare ODD absurd and view this result as a refutation of Lewis's strategy. *Strong* is supposed to be the natural understanding of ability in the premises of CA for incompatibilists — "whereby they mean the ability to bring something about or cause something to be by virtue of one's actions" (Kane 1996: 50). With premise (4) impeccable, the rejection of ODD and a simple *modus tollens* re-establish the thesis of incompatibilism: given determinism, it is not the case that one can otherwise.

The MCR agrees with the negative verdict on ODD and proposes a different strategy. Once the distinction weak/strong is available, why not apply it to premise (4) of CA_1?

(4') If J could have rendered *P* false, and if the conjunction of *H* and *L* entails *P*, then he was able to do something such that, if he did it, the conjunction would have been false.[8]

While J is able to render *P* false in the strong sense, J is not able to render this conjunction false in the strong (causal) sense; he cannot bring about the difference. According to Perry (2004: 247), when we interpret "can render false" in the strong, causal sense (the proposition is false *because* of my action), then (to put *P* instead of his *Q*): "It simply does not follow from the fact that J will render *P* false that he renders false every proposition that entails *P*. What does follow is that there is no true proposition that entails *P*."

Kapitan (2011: 137–138) suggests the same move when he considers the case of a soldier (Roni) who is able to disobey his commander's instructions, but nevertheless decides to act as commanded ('P'). He claims that Roni is *broadly* able to bring about '~(H & L)', but not causally:

> If /.../ ability is construed in the causal sense, then the premises of the Consequence Argument are true; /.../ there is nothing Roni is able to do that would make it the case that either *H* or *L* does not obtain, and so, Roni is causally unable to see to it that ~(H & L). But, in that case, because Roni is able to disobey orders and by so doing would make it the case that ~P, he is causally able to see to it that ~P. Consequently, the transfer rules for causal unavoidability are invalid. Likewise, because Roni is causally able to see to it that ~P and ~(H & L) is a consequence of ~P, then the transfer rules for causal ability are invalid.

As the last sentence indicates, Kapitan clearly thinks that Beta 2 (closure under *entailment*) fails if we adopt a causal (strong) reading of (in) ability. Roni is causally *unable* to see to it that ~(H & L). The defenders of MCR in general agree that the transfer principle behind premise (4) is false. They propose to deny premise (4) with the help of *broad* and thus avoid ODD. A free predetermined agent can act otherwise in a strong ("causal") way but can falsify the conjunction in the *weak* (broad) sense only. Her action at time t will not make the conjunction untrue, it will not render it untrue. Instead, the conjunction will "be untrue" because the earlier events at time t–1 made it false (Perry, 2004: 249, see also Rummens, 2019). From the logical point of view, MCR declares both critical principles, Beta (closure under *implication*) and Beta 2 (closure under *entailment*), fallacious. But, in denying the latter, they disagree with Lewis.

One might propose a reading that makes ODD more acceptable to the wider compatibilist camp. Being able in a *strong* sense to falsify the conjunction of history and law is simply to be able to cause an event

[8] To use the "simply" weak sense of "can render false", preferred by Lewis instead of "weak" in his letter to Horgan (Lewis et al. 2020: 119).

E such that the conjunction is false in every possible world in which E occurs. And the claim that deterministic agents possess this ability is something that any compatibilist who believes in deterministic agents' ability to sometimes act otherwise, needs to accept. But this interpretation blurs the very distinction between weak and strong and it goes against the textual evidence. Consider Lewis again:

> Let us say that I could have rendered a proposition false in the weak sense iff I was able to do something such that, if I did it, the proposition would have been falsified (though not necessarily by my act, or by any event caused by my act). And let us say that I could have rendered a proposition false in the strong sense iff I was able to do something such that, if I did it, the proposition would have been falsified either by my act itself or by some event caused by my act (1986: 297).

This passage strongly suggests that, understood in the strong sense, my act itself (or something caused by my act) would *render* the conjunction of history and law untrue. The proposed reading of *strong*: "being able in a strong sense to falsify P is to be able to cause an event E such that P is false in every possible world in which E occurs" is suspiciously close to *weak* (just being able to do something such that, if S did it, P would have been falsified). Is it possible to defend a view that according to which *strong* ability with respect to a certain true proposition can be interpreted so that: (i) the proposition would have been falsified either by my act itself or by some event caused by my act; (ii) the proposition would not have been thereby *rendered* false either by my act itself or by some event caused by my act; (iii) strong does not collapse into weak; (iv) and (i) still ascribes to the agent more than just the ability to do something such that, if she did it, the proposition would have been (weakly) falsified? Difficult to say, it is not easy to resolve all the tensions generated by Lewis's distinctions. Let me try to address these issues from the general point of the ability to render a proposition false.

6. *Loaded and neutral*

Recall the general notion of ability to render a proposition false (Kapitan 2011: 135): S is able at t to see to it that $\sim p$ iff there is a course of action X such that at t (i) S is able at t to do X, and (ii) were S to do X, then $\sim p$. One way of understanding the second condition is *broad*; in most general terms, S's action would be inconsistent with the truth of p, or, from S's doing X, it may be inferred that p is false. Even logical necessities and past truths are such that one is able at t to see to it that they obtain in this sense: they are (now) true and whatever S does they (still) remain true. Let 'P' stand for my actual refraining to raise my hand, although I am perfectly able to do so. Let 'Q' be a contradiction, and 'R' an arbitrary false proposition about the past (say: "On July 20, 1969 Edwin Aldrin became the first human to step on the Moon"). The closure of the ability operator 'A' under entailment licenses the inferences: (i) "A~P, Q entails P ⊢ A~Q" and (ii) "A~P, (P & R) entails P ⊢ A~ (P & R)."

According to (i), I am able to render a contradiction false (and so a tautology true). For any course of action that I can take at *t*, it would (still) be the case that not–Q. According to (ii), I am able to render false a conjunction which was *already* false independently of my actions (Neil Armstrong was the first human to step on the Moon). For any course of action *P* that I can now take (after, say 2020), it would (still) be the case that not–R; thus, the conjunction 'P & R' would remain false.

This is the "neutral" interpretation of ability sometimes championed by modal logicians. However, this neutral, purely logical (modal) notion comes with a price: it may sound unnatural and contra-intuitive. Even some logically minded people will protest, thus Kenny (1976: 214):

> The President of the United States has the power to destroy Moscow, i.e., to bring it about that Moscow is destroyed; but he does not have the power to bring it about that either Moscow is destroyed or Moscow is not destroyed. /. . ./ The power to bring it about that either *p* or not *p* is one which philosophers, with the exception of Descartes, have denied even to God.

According to Kenny, nobody has the power to bring it about that a tautology is true, which seems equivalent to the ability to render a contradiction false. In a similar way, Perry (2004: 247) objects to (ii); nobody can render false a past falsity (the gist of his objection to premise (4) of CA_1). Contradictions and past falsities are not *made* false by the agent; they are just false independently of their (present) intentions and interventions. In a similar vein, Schneider (2004: 418) defines van Inwagen's "x can render *p* false" as "x can do something such that if x did it, *because* of that *p* would be false," treating "because" as expressing a primitive "explanatory relation." Premise (4) of CA_1 fails on this reading. Nothing that the judge can do is such that *because* of his doing so, the conjunction of *L* and *H* would be false; the conjunction is false for different reasons (although Schnieder does not say what those reasons are (Schnieder 2004: 423).

I will call the notion of ability suggested by Kenny, Perry and Schnieder "loaded" ability, roughly associated with the "active", causal power to bring something about, realize, change something, or make something false. The active component can be further strengthened with explanatory and/or agential components, perhaps cognitive and intentional. Active "causal" contribution is enough to block closure under entailment, but when we include additional constraints imposed by the general "metaphysics of agency", the invalidity of closure under entailment becomes even more apparent. To take an example by Kapitan (1996: 423): I am able to drink a cup of coffee. By so doing, I would bring about complex molecular changes in my brain, but I am ignorant of physiology, so these changes are not a reliable result of my drinking coffee. I am not able to bring them about, so closure fails.

To summarize—suppose we take "loaded" ability as an umbrella term for the notion of ability involving active, causal powers (by doing X, the agent causally contributes to the obtaining of *p* or makes

it the case that *p*). We might impose further constraints (reliability, intentionality, knowledge, skill, etc.) and thereby generate layers of practical modalities (cf. Kapitan 1996). According to MCR, causal ability in CA should be interpreted as loaded, but closure under entailment *fails* for loaded, so premise (4) of CA_1 is false. Broad ability, however, is neutral from the logical point of view. One might perhaps introduce different levels of neutrality. On a higher level, there is the *general* rule of consequence (the generalization of rule Beta or closure under implication); on a lower level, there is *simple* consequence (closure under entailment—*Master* or Beta 2, see Appendix).

Lewis denies the higher level of neutrality for *strong* ability, but he accepts the lower. This is a signal that Lewis's strong/weak distinction does not map unproblematically onto the causal/broad distinction of MCR. There are two options: (1) strong ability is not necessarily causal, or (2) the logic of strong ability is not the logic of loaded ability. One can test the first option: my act, such that I was predetermined not to perform, would itself be an event directly and reliably falsifying the broad past without thereby *making* any changes. Perhaps one "can directly bring about circumstances in which *p* is false", but "one cannot bring about the falsity of *p*" (Brown 1988: 24, fn 13). This "direct" but non-causal ability will still be stronger than "indirect" weak ability. This interpretation might further be supported by Lewis's views on causation and counterfactual dependency. By his account, causal dependence requires the appropriate patterns of counterfactual dependence among two *particular* events: *C* and *E* (*E* causally depends on *C* if and only if: if *C* had occurred, *E* would have occurred; and if *C* had not occurred, *E* would not have occurred). J can then cause the conjunction of *H* and *L* to be different only if there is some *particular* event that counterfactually depends on J's raising his hand. Nevertheless, differences in the conjunction of *H* and *L*, correlated with J's counterfactual raising of his hand, might have been realized in a variety of ways.

Still, we might be inventing distinctions where there are none. Take Lewis's letter to Lehrer where he says that they agree in distinguishing the following (Lewis et al. 2020: 94):

(1) I could have done so-and-so, and if I had, the laws (or history) would have been different,
(2) I could have brought it about that the laws (or history) would have been different.

For both for Lewis and Lehrer, (1) is true and (2) is false. According to Lewis, the second statement adds some false extra content that has to do with *causation* from my action to the law-breaking miracle (or the alternative past). The extra content, apparently characterizing *strong* ability, is causal. Moreover, Lewis's examples (Lewis 1981) and comments suggest a causal reading for strong (ability to break the window, ability to break a promise). Although some tensions will remain, the second option is more plausible: strong is causal, but the logic of strong

is not the logic of loaded. I conjecture that Lewis follows the practice of (some) logicians who accept the closure of ability under entailment with all the corollaries alluded to earlier: an agent is able to bring it about the logically unavoidable. At least in pre-*stit* logical practice, it was accepted as technically convenient and unproblematic to construe the *ability* operators as meaningfully applicable even to necessarily true sentences (Brown, 1988: 24). It is not impossible to defend such a practice; consider the comment by Chellas on Belnap, whose notion of "the agent sees to it that …" is *loaded* (not closed under logical consequence):

> Can it ever be the case that someone sees to it that something logically true is so? I believe the answer is yes. When one sees to something, one sees to anything that logically follows, including the easiest such things, such as those represented by a tautology. One should think of seeing to it that (e.g.) 0 = 0 as a sort of trivial pursuit, attendant upon seeing to anything at all. (Chellas 1992: 508)

Certain remarks made by Lewis in a different context suggest that he would accept this line of reasoning. For instance, when discussing the logic of relevance and its motivation (supposedly problematic inferences *ex falso quod libet* and *verum ex quod libet*), he explains tautology as vacuously about any subject matter and thus, one might assume, also trivially implied by and a *result* of any action the agent is able to perform (Lewis 1988: 115). There is a price to be paid as a counterexample to loaded show; it is difficult to see how logical necessities are "made true" by the agent. Nevertheless, there are also benefits—above all, generality and liability to logical investigations. Extensionality is no longer a prerequisite; van Inwagen, when faced with the properties of *loaded* ability as an objection to his official definition of "can render false", claimed that his definition at least had the consequence that "S can render *p* false" was a purely extensional context (van Inwagen 1983: 231, fn 9). In the contemporary framework of modal logic, extensionality is not sacrosanct. Even "normality" can be sacrificed (a generalization of Beta or closure under implication is required for "normal" modal logic). However, an operator that is not closed under logical consequence satisfies virtually *no* logical laws.

7. *Odd?*

I have argued that Lewis's way with closure and premise 4 of CA_1 is the way of standard (pre-*stit*) logicians on ability. Admittedly, the defenders of loaded have made some persuasive points; logical laws are not sacred, and the tools of logics have evolved (in the framework of contemporary *stit* approaches, the ability is not closed under logical consequence). Lewis's strong ability is causal but nevertheless neutral from the logical point of view. He might or might not agree with some of the proposed tweaks, but he never questioned *Master* (see a discussion about *Hasker* in the *Appendix*). In his letter to van Inwagen, he acknowledges that a compatibilist might worry that by denying prem-

ise (5), "the compatibilist has a problem explaining why he wouldn't be reversedly causing a divergence miracle" (Lewis et al. 2020: 91). By denying "strong", he solves the problem: J could have rendered antecedent, the conjunction of *H* and *L* false (in the strong, causal sense), but J could not have rendered the consequent *L* false in the strong sense. However, the compatibilist does not face the problem of explaining why he wouldn't be reversedly *causing* a difference in the conjunction of *H* and *L*. Why not?

Setting aside Lewis's somehow idiosyncratic views on causation and counterfactuals, there is a way to explain and even defend this move. Some new textual evidence comes from his recently published correspondence and previously unpublished manuscripts. Lewis in his letter to Thomas Nagel, remarks that, "I could have raised my hand" is really inconsistent with history and the law, but compatibilists should not bother, since "It's scarcely even a consequence of compatibilism — just a restatement of it" (Lewis et al. 2020: 94). I propose to extend this type of reasoning to strong ability with respect to the conjunction of *H* and *L* in CA_1. Strong ability to render false this conjunction is scarcely a consequence of the ability to do otherwise, given the truth of determinism: it is merely a restatement of it.

Horgan hinted at this line when he introduced a variation on *strong*. Under this interpretation, J can raise his hand at T and this act *itself* is now being counted as a *H* (or *L*)-falsifying event. But, according to Horgan:

> It would be outrageous, of course, to claim that J can causally influence events in the remote past. But we are saying nothing so offensive when we assert that J can render *H* false in the strong and broad sense. On the contrary, essentially all we are saying is that J can do something that he is causally determined not to do; and it is no surprise to learn that the compatibilist is committed to *that*. (1985: 348)

In the opposite camp, Kane, an incompatibilist, noted the same fact and, of course, duly objected to *that*. In his version of CA_1, the crucial premise (4) is renumbered as premise (3), while premise (6) reads as a denial of ODD: "It is not possible that an agent *a* at *t* can (has the power or ability to) render false the conjunction of distant past and law." In his discussion of the compatibilist opposition to *Master* (premise (3) in his reading), he writes (Kane 1996: 51):

> Premise 3 says 'if (i) you could have done other than move your hand, and (ii) your hand's moving was determined by laws and the past, then you could have rendered false a law of nature or the past' (which is the denial of 6). But to us incompatibilists, assuming at the outset that (i) 'you could have done other than move your hand' under the assumption that (ii) 'your hand's moving was determined' begs the whole question. For it means that the power to do other than move your hand that is assumed in the antecedent of 3 must be a compatibilist power—which means in turn that the argument against 3 succeeds only if one assumes at the outset a compatibilist interpretation of the power to do otherwise.

Clearly, he is targeting the position of MCR on strong (and not Lewis)—if you reject strong ability to render false the conjunction of *H* and *L* you have to object to *Master* (closure under entailment), and he finds *any* objection to Master question-begging. The very idea of a counterexample to Master is problematic; for the conditional (his premise 3, premise (4) of the original CA_1) to be false, the antecedent must be true and the consequent false. However, in the antecedent, we begin with the *assumption* that J can do something that he is causally determined not to do, which is already question-begging according to Kane. Lewis, as we saw, does not reject strong ability to render false the *conjunction* of history and law and finds no problem with *Master*. Nevertheless, he is also committed to the compatibilist interpretation of the power to do otherwise, according to Kane (1996: 223 fn 12):

> We might easily overlook this fact because the argument assumes Lewis's strong sense of "can do otherwise," which is not so obviously a compatibilist notion as is his weak sense. But, as Lewis is well aware, his strong (or causal) sense of "can do otherwise" is not necessarily incompatibilist. It can also be given a compatibilist analysis.

Let us take stock of the debate and various positions. ASSUMPTION will be the claim that a free predetermined agent can act otherwise. Given determinism and suitable transfer principles, (*Master*) ASSUMPTION implies ODD, the ability to render false the conjunction of distant past and laws of nature. Let us call WONDER the possession of *strong* ability to render false *one* of the conjuncts of ODD. Let us call PECULIAR the possession of *weak* ability to render false the conjunction of distant past and laws of nature. Finally, let us call MIRACLE the possession of *weak* ability to render false *one* of the conjuncts of ODD (laws of nature, according to Lewis's brand of local miracle compatibilism).

Incompatibilists take *Master* as beyond any doubt, but ODD is already incredible enough (Kane 1996: 50–52), so the ASSUMPTION must go. Some compatibilists (the MCR group) agree that ODD is incredible and WONDER must be avoided at all costs. Therefore, *Master* must be denied, but PECULIAR is something a compatibilist can live with. Lewis is unique: he sides with the incompatibilists in accepting *Master*, embraces ODD as a result but denies that WONDER follows from ODD. According to Lewis, ODD implies MIRACLE (with respect to the laws of nature) only, since Agglomeration fails (strong ability to render false a conjunction does not distribute over the conjuncts). Moreover, according to Lewis, both MIRACLE and ODD are something a compatibilist can live with.

This is just a skeleton; there are more nuanced positions (cf. van Inwagen: 2004, 349), and I did not discuss how to live with MIRACLE in this paper. I mainly agree with Vihvelin (2013: 164) that the miraculousness is just a result of our counterfactual speculation about what would have been the case, beforehand, if anything in a determin-

istic world had happened in any way other than the way it actually happened. If counterfactuals in a deterministic world make any sense, then so does MIRACLE. However, my main concern is how to live with ODD-ness according to Lewis. I will suggest that the acceptance of ODD can be interpreted as a corollary of a *Moorean* fact that we have free will. The compatibilists who are embarrassed by ODD underestimate the *dialectical* force of "strong."

8. *Moorean facts*

Many of Lewis's philosophical investigations start from not negotiable pieces of our ordinary picture of the world (cf. Nolan 2015). This is also true of his compatibilism (Lewis 2020: 241):

> It's a Moorean fact that we often have a choice what to do. But whether determinism holds is an unsettled question. So, having a free choice is epistemically compatible with determinism. And with indeterminism. So, it's compatible simpliciter.

We begin with a Moorean fact that we are able es to act otherwise, and an arbitrary person on the street, ignorant of philosophical technicalities, will very likely understand their ordinary ability to raise their unraised arm in causal terms. But then being able (in the ordinary, Moorean sense) to act otherwise if determinism is true amounts to being able to falsify the conjunction of history and laws of nature that determines the actual action. Consider an analogy. According to scientific essentialism, if gold exists, then it has—necessarily—atomic number 79. Possessing this property is a metaphysically necessary condition for being the kind of thing designated by the natural kind term "gold". Suppose, somehow anachronistically, that two alchemists (A1 and A2) are discussing the following statement made by A1:

PS If scientific essentialism is true and by using a philosophers' stone you are able to turn this piece of stuff which is iron into gold, then you are able to change the atomic number of this stuff from 26 to 79.

A2 might object that it is impossible to change the atomic number of an element. But A1 will reply, that, given scientific essentialism, that is just what it *means* to turn iron into gold. By turning iron into gold, you are changing the atomic number of this stuff. Of course, A2 might object to the doctrine of scientific essentialism on its own, but this is a separate issue and not an objection to PS. And A2 might think that the truth of scientific essentialism (somehow?) precludes the ability to effectively use the philosopher's stone. But then he would still agree with PS. Given scientific essentialism, the act of turning this piece of stuff, which is iron, into gold would be truly describable as changing the atomic number of this stuff from 26 to 79. And given determinism, the act of raising my actually unraised hand would be truly describable as rendering false the conjunction of distant past and the laws of na-

ture, to use the terminology of Fischer in his version of the consequence argument (Fischer 1994: 28).

Let us call sentences of the type "X would be truly describable as Y", where the truth of the proposition expressed is grounded in the entailment relation between the contents of X and Y, "quasi-analytic." It is then quasi-analytically true that being able to act otherwise given that determinism is true is just being able to falsify the conjunction of history and the laws of nature which entails the actual action. *Master* is then almost trivial—exactly as van Inwagen (1983: 72) claimed it to be. And to use the phraseology of Chellas on Belnap, one should think of seeing to it that the conjunction is false as a sort of pursuit, attendant upon seeing it to do what one was predetermined not to do. Looking for counter-examples is then almost futile; this will disturb the MCR branch of compatibilism. This will also disturb those incompatibilists who use *Master* as their main principle for deriving *consequences* inimical to compatibilism (Finch and Warfield 1998).

According to this line of reasoning, ODD is a quasi-analytic restatement of compatibilism. This is just what it means to be predetermined but still be able act freely. When ascribing to J strong ability to render false the conjunction of history and law, then essentially all we are saying is that J can do something that he is causally determined not to do; and it is no surprise to learn that the compatibilist is committed to *that*. Kane is perhaps aware of this fact when he argues that a denial of *Master* is already question-begging: not because some magical powers are assigned to the agents but because of the very ASSUMPTION (compatibilism). Of course, to reject an account of the ability to do otherwise on the sole ground that it is compatible with determinism begs the question just as well.

9. *The limits of logic*

I have explored some rather intricate details of CA and practical modalities involved (cf. also the *Appendix*). I think that logical investigations help us to systematically extract the consequences of our initial commitments. But the philosophical role of an argument and its validity is overrated by Murdoch (the initial quotation). I think that Harman is ultimately right: inference is always "inference to the best overall view." The acceptability of the premises and our starting points are an important ingredient of our overall view. Well, Lewis is unusually clear about *his* starting points. For instance:

> Apart from that [believing in the existence of concrete alternative possible worlds] I am philosophically conservative: I think philosophy cannot credibly challenge either the positive convictions of common sense or the established theses of natural sciences and mathematics. (Pyke 1995)

It is a firm conviction of common sense, a Moorean fact, that we make free choices. So compatibilism and libertarianism (free will being in-

compatible with determinism) are the only philosophical options left. I think that the respect for science is then decisive for Lewis: "we know better that we are sometimes free than that we ever escape predetermination; wherefore it may be for all we know that we are free but predetermined" (Lewis 1993: 155). For all we know, the thesis of determinism might be true, but it is up to science to (dis)confirm this thesis. A weak/strong distinction then still allows for two compatibilist options: to accept or to deny closure under entailment for strong ability.

At this point it might be useful to compare Lewis's position with respect to strong ability to act otherwise (the acceptance of closure) with a familiar Moorean position on knowledge and radical skepticism. The logical parallel of *Master* is the claim that knowledge is closed under entailment. Roughly, "If S knows that *P*, and *P* entails *Q*, then S knows that *Q*." This must be refined[9], but it will suffice for our comparison. Knowledge closure figures prominently in a much-discussed argument for skepticism. Our ordinary perceptual knowledge logically excludes radical skeptical scenarios where these scenarios are subjectively indistinguishable from a paradigm case of perception, but where one is in fact massively deceived. The radical skeptic then uses closure in the following familiar argument:

> I know I am standing; my knowing that I am standing entails that I am not dreaming; but I do not know that I am not dreaming, so I do not know that I am standing.

Moore is famous for agreeing with the sceptic in *accepting* the relevant closure principle but not giving up our ordinary knowledge; he thus argues:

> I know I am standing; my knowing that I am standing entails that I am not dreaming; therefore, I know that I am not dreaming.

Consider now the beginning of Lewis (1986: 291), paraphrased in terms of closure:

> I can raise my hand. Given the truth of determinism my doing so entails that I am able, in a strong sense, to render false the conjunction of *H* and *L*. Therefore I am able, in a strong sense, to render false the conjunction of *H* and *L*.

I know I am standing and the way in which I know is not the sort of way that is endangered by the possibility of a radical, skeptical scenario according to Moore. And I know I am free. The way in which I am (perhaps) determined not to do so is not the sort of way that counts as inability according to Lewis. A classical Moorean strategy allows one to meet the challenge of skepticism without having to deny the closure principle. Moreover, a parallel Moorean strategy allows one to meet the challenge of incompatibilism without having to deny the closure principle for (strong) ability.

[9] For instance: "If S knows that *p*, and S competently deduces from *p* that *q*, thereby forming a belief that *q* on this basis while retaining their knowledge that *p*, then S knows that *q*" (Pritchard 2016: 13).

Both a Moorean position on knowledge and a Moorean attitude on free choice have seemingly incredible consequences. Nevertheless, the cases are not quite analogous from the *argumentative* point of view. One of the crucial elements in the Moorean anti-skeptical strategy is a plausible explanation of our knowledge of the denials of skeptical hypotheses. Moore's comments are often found puzzling (Moore 1993: 169): "I have, no doubt, conclusive reasons for asserting that I am not now dreaming; I have conclusive evidence that I am awake: but that is a very different thing from being able to prove it." It is notoriously difficult to account for this evidence. (Neo)Mooreans in general endorse our common-sensical knowledge of the denial of the radical skeptical hypothesis, but few are willing to claim that the entailment "I am standing entails that I am not dreaming" is capable of *transferring* such knowledge.

Lewis never bothered with addressing the analogous problem of explaining our strong ability to render false the conjunction of history and law. I think there is an important difference between the two cases which might explain this "carelessness." Moore must provide some reasons for the claim that we know a radical skeptical hypothesis to be false. Lewis, on the other hand, must explain our strong ability with respect to the *conjunction* of *L* and *H*. According to my interpretation, ODD is quasi-analytic, merely a restatement of a compatibilist position with respect to ability to act otherwise. To use the phraseology of Chellas on Belnap, one should think of seeing to it that the conjunction is false as a sort of pursuit, attendant upon seeing it to do what one was predetermined not to do. There are no marvelous powers at issue, nothing is literally "transferred." Explaining the evidence for our (in)ability with respect to *each* of the conjuncts (A~H, A~L) and how those (in) abilities do or do not combine (Agglomeration) is where all the action is. From the purely evidential point of view, the reasons for thinking that the past is fixed for an agent are different from the reasons for thinking that the laws are fixed. The past (or distant past) is commonly understood as inevitable ("Everything that is past and true is necessary", according to Diodorus Cronus). However, the laws of nature might be Humean: as facts about regularities among events, they predict but do not constrain. In any case, van Inwagen is aware of the dialectics and carefully provides evidence for each premise of the consequence argument ('NH' and 'NL') separately. The weak/strong distinction is an attempt to refute this evidence.

And finally, to answer my original question: if compatibilism is the *accused*, whom should we praise—defense or prosecution with its CA? Well, who bears the burden of proof and who has the presumption of innocence? A Moorean stance combined with the respect for science takes compatibilism as the initial (not guilty) position. In this case, we should agree with Lewis that the prosecution with its consequence argument failed to prove the case. Still, at the end of the day, this is not a purely "logical" victory. Ramsey (initial quotation) might be right, logic just

displays the structure of your initial commitments, when you disagree with certain conclusions, you disagree with a certain set of starting points (thus you "beg the question"). But it shows this much: a certain set of starting points does not entail an absurd conclusion. Not really what Murdoch expected from a philosophical argument, but perhaps all that can be realistically expected as a *probative* role of logical tools in the case of the consequence argument.

Appendix

Lewis (1981) discusses the first version of the consequence argument where "can(not) render false", expressed here in terms of ability ('A'), is the crucial notion. Most contemporary discussions focus on the third, modal version, where the "no choice" ('N') necessity is central. In order to compare them, we must establish a translation scheme between the two notions and compare some central principles:

Alpha	$\Box p \vdash Np$
Agglomeration	$Np \,\&\, Nq \vdash N(p \,\&\, q)$
Beta	$Np, N(p \rightarrow q \vdash Nq$
Beta 2	$Np, \Box(p \rightarrow q) \vdash Nq$
Master	$A{\sim}q, \Box(p \rightarrow q \vdash A{\sim}p$

I rely, roughly, on Kapitan's proposal (2011: 131) and interpret "S can render *p* false" as: "S is able prevent that a situation *p* obtains." Having no choice about *p* or 'Np' is then naturally defined as "p & ~A~p" or "*p* is true and nobody is able to prevent *p* from being true." Caution is required, however. Tautologies are logically necessary truths, so it follows (via rule Alpha), that no one has, or ever had, any choice about whether a tautology is true. But let 'R' stand for the proposition that expresses the fact that J did not raise his hand at T and let 'Q' denote an arbitrary contradiction. We assume that J could have rendered *R* false, *Q* entails *R* (a contradiction entails anything), so, via *Master*, J could have rendered *Q* false?! This result follows even from van Inwagen's official definition of "S can render *p* false" (van Inwagen 1983: 68):

> It is within S's power to arrange or modify the concrete objects that constitute his environment in some way such that it is not possible in the broadly logical sense that he arrange or modify those objects in that way and the past have been exactly as it in fact was and *p* be true.

Well, whatever S does—it is impossible for S to arrange or modify their environment and a contradiction be true (cf. Schnieder 2008: 106). Let us take an arbitrary contradiction 'Q' and its negation '~Q' which is, assuming classical logic, a tautology. According to the translation schemes proposed *Master* licenses a strange result. No one has, or ever had, any choice about whether a tautology is true ('N~Q'), but one *can* prevent a contradiction from being true, which seems equivalent to being able to see to it that the *very* tautology in question ('A~Q') obtains. This looks strongly counterintuitive.

However, let me first note that 'Np' is defined as a *conjunction*: *p* is true *and* nobody can prevent *p* from being true. We get a better match between the "no choice" of CA_3 and "can(not) render false" of CA_1 when we include the first conjunct in the closure principle governing 'A' (*Hasker*[10] instead of *Master*):

$$A{\sim}q, \Box(p \rightarrow q), p \vdash A{\sim}p$$

The principle restricts the scope of one's ability to the ability to prevent a (contingently) true proposition from being true. Van Inwagen himself hinted at such a restriction when discussing certain odd consequences of his official definition of the ability to render a proposition false (S can render false an arbitrary falsity about the past). However, the first argument for the incompatibility of free will and determinism involves *true* propositions only (van Inwagen 1983: 68).

Different understandings of "ability" resonated importantly in our discussion of *Master*. The ability to see to it that a tautology obtains is strange when this ability is associated with "active", causal power to bring something about. But "loaded" (causal, active) modal notions satisfy virtually no logical laws, so there is not much to say about the principles governing them. I have argued that Lewis would accept the claim that one is *able* to render a contradiction false. The conjunction "N~Q & A~Q" is odd but not "abominable." When both locutions are translated in terms of neutral 'A' we get an asymmetry. One is unable to *prevent* a tautology (which is, of course, true) from being true, but one is *able* so to act that a contradiction is false.

Given these explanations, we can show that *Master* (and a revised *Hasker*) implies Beta 2. I will assume standard propositional logic (PL) and the fact that broad logical necessity is *alethic* (also compare Kapitan 1991: 335):

1. Np	hypothesis
2. $\Box (p \rightarrow q)$	hypothesis
3. $\neg A \neg p \ \& \ p$	1 definition of 'N'
4. $A\neg q \rightarrow A\neg p$	2 Master
5. $\neg A \neg p \rightarrow \neg A \neg q$	4 PL
6. $\neg A \neg p$	3 PL
7. $\neg A \neg q$	5, 6 PL
8. $p \rightarrow q$	2 modal logic
9. p	3 PL
10. q	9, 8 PL
11. $\neg A \neg q \ \& \ q$	10, 7 PL
12. Nq	11 definition of 'N'

[10] Hasker (1989: 112) defends the principle (PEP 5): If it is in S's power to bring it about that P, and "P" entails "Q" and "Q" is false, then it is in S's power to bring it about that Q. (PEP 5) is equivalent to *Hasker*.

What is the relation between various closure principles (Master, Beta, Beta 2)? Beta 2 and Agglomeration together entail the validity of Beta, just for the record:[11]

1. N p	Premise
2. N (p → q)	Premise
3. N [p & (p → q)]	1, 2 Agglomeration
4. □ {[p & (p → q)] → q}	Necessity of a logical truth
5. N q	3, 4 Beta 2

A counter-example to Beta is therefore a counter-example to the combination of Agglomeration and Beta 2. One might accept Beta 2 as valid but still deny Beta because Agglomeration fails for the relevant operator; this is Lewis's position, according to my interpretation. Lewis accepts premise (4) of CA_1, an instance of *Master*, and we have established that this principle implies Beta 2. He explicitly denies Agglomeration for the dual of *strong* ability, and we also know that he challenges Agglomeration for the "no choice" operator in general.

Lewis (1993) discusses the operator "it is true that, and such-and-such agent never had any choice about whether …" abbreviated as "Unfree". According to Lewis, the best argument for incompatibilism (CA, apparently) rests on a plausible principle that "Unfree" is closed under implication. Suppose that some premises imply a conclusion, and we prefix "Unfree" to each premise and to the conclusion. According to the closure principle, the prefixed premises imply the prefixed conclusion. Here is the full text (Lewis 1993: 169–170, fn 11):

> The closure principle is a generalization of the 'Rule Beta' that plays a leading role in Peter van Inwagen's defence of incompatibilism in *An Essay on Free Will* (1983); it first appears on page 94. The closure principle says that the logic of 'Unfree' is a 'normal' modal logic, see Brian Chellas (1980: 114–115). We can see from Chellas's Theorem 4.3(4) that the closure principle is equivalent, *inter alia*, to this combination of four principles:
>
> RE: if 'A if B' is valid, so is 'Unfree A if Unfree B',
> N: 'Unfree T' is valid, where T is an arbitrary tautology,
> M: 'Unfree (A & B)' implies 'Unfree A and Unfree B,' and
> C: 'Unfree A and Unfree B imply 'Unfree (A & B)'.
>
> The compatibilist must therefore challenge one of the four, most likely C, and Michael Slote (1982) has done so.

Let us embed this remark in the more general framework of modal logic. Let 'O' stand for an appropriate modal operator ("(in)ability", "(un) avoidability", etc.). The general logical principles governing this operator are then (where 'T' is tautology):

[11] Carlson (2000: 288, fn 12) gives credit to Krister Bykvist for this derivation. However, Chellas (1980: 122) in his 4.5.b already indicates this result.

RE.	From $\vdash A \leftrightarrow B$ infer $\vdash OA \leftrightarrow OB$
N.	OT
M.	$O(A \& B) \rightarrow (OA \& OB)$
C.	$(OA \& OB) \rightarrow O(A \& B)$

The four principles jointly result in the *general* rule of consequence:

RK. From $\vdash (A_1 \& A_2 \& \ldots A_n) \rightarrow A$ infer $\vdash (OA_1 \& OA_2 \& \ldots OA_n) \rightarrow OA$

Beta is then just a special case of RK for 'N':

From $\vdash (A \& (A \rightarrow B)) \rightarrow B$ infer $\vdash (NA \& N(A \rightarrow B)) \rightarrow NB$

Rule RE together with M yields:

RM. From $\vdash A \rightarrow B$ infer $OA \rightarrow OB$

This rule expresses a weak or *simple* consequence: if *B* is a consequence of *A*, then OB is a consequence of OA. We can easily obtain *Master* from RM ("p entails q" is equivalent to "~q entails ~p", so, according to RM, 'A~p' is a consequence of 'A~q').

References

Brown, M. 1988. "On the Logic of Ability." *Journal of Philosophical Logic* 17: 1–26.

Carlson, E. 2000. "Incompatibilism and the Transfer of Power Necessity." *Noûs* 34: 277–290.

Chellas, B. F. 1980. *Modal Logic: An Introduction*. Cambridge: Cambridge University Press.

Chellas, B. F. 1992. "Time and Modality in the Logic of Agency." *Studia Logica* 51: 485–517.

Finch, A. and Warfield, T. A. 1998. "The *Mind* argument and libertarianism." *Mind* 107: 515–28.

Fischer, J. M. 1994. *The Metaphysics of Free Will*. Oxford: Basil Blackwell.

Hasker, W. 1989. *God, Time and Knowledge*. Ithaca: Cornell University Press

Horgan, T. 1985. "Compatibilism and the Consequence Argument." *Philosophical Studies* 47: 339–356.

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

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

Kane, R. (ed.). 2002. *The Oxford Handbook of Free Will* 1st. Oxford: Oxford University Press.

Kane, R. (ed.). 2011. *The Oxford Handbook of Free Will* 2nd. Oxford, New York: Oxford University Press.

Kapitan, T. 1991. "How Powerful Are We?" *American Philosophical Quarterly* 28: 331–338.

Kapitan, T. 1996. "Modal Principles in the Metaphysics of Free Will." *Noûs* 30: 419–445.

Kapitan, T. 2002. "A Master Argument for Incompatibilism?" In Kane, R. (ed.). 2002: 127–157.

Kapitan, T. 2011. "A Compatibilist Reply to the Consequence Argument." In Kane, R. (ed.). 2011: 131–150.

Kenny, A. 1976. "Human Abilities and Dynamic Modalities." In Manninen, J. and Tuomela, R. (eds.). *Essays on Explanation and Understanding*. Springer: 209–232.

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

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

Lewis, D. 1988. "Relevant Implication." *Theoria* 54: 161–174.

Lewis, D. 1993. "Evil for Freedom's Sake?" *Philosophical Papers* 22: 149–172.

Lewis, D. K. 2020. "Outline of 'Nihil Obstat: An Analysis of Ability.'" *The Monist* 103: 241–44.

Lewis, D. K., Beebee, H. and Fisher, A. R. J. (eds.) 2020. *Philosophical Letters of David K. Lewis: Volume 1: Causation, Modality, Ontology*. Oxford, New York: Oxford University Press.

McKay, T. and Johnson, D. 1996. "A Reconsideration of an Argument against Compatibilism." *Philosophical Topics* 24: 113–22.

Mele, A. R. 2006. *Free Will and Luck*. Oxford: Oxford University Press.

Nolan, D. 2015. "Lewis's Philosophical Method" In Loewer, B., & Schaffer, J. (eds.). *A Companion to David Lewis*. John Wiley & Sons: 25–39.

Perry, J. 2004. "Compatibilist options." In Campbell, J. K., O'Rourke, M. and Shier, D. (eds.). *Freedom and Determinism*. Cambridge: MIT Press: 231–254.

Pyke, S. 1995. *Philosophers*. London: Zelda Cheatle Press.

Pritchard, D. 2016. *Epistemic angst: Radical skepticism and the groundlessness of our believing*. Princeton: Princeton University Press.

Ramsey, F. P. 1994. *Philosophical Papers*. Cambridge: Cambridge University Press.

Rummens, S. 2019. "The Counterfactual Structure of the Consequence Argument." *Erkenntnis*, https://doi.org/10.1007/s10670-019-00117-2.

Schnieder, B. 2008. "On what we can ensure." *Synthese* 162: 101–115.

Setiya, K. 2020. "Philosophy as Self-expression." *The Philosopher* 108 (3): 65–69.

Slote, M. 1982. "Selective Necessity and the Free-Will Problem." *The Journal of Philosophy* 79: 5–24.

Speak, D. 2011. "The Consequence Argument Revisited." In Kane, R. (ed.). 2011: 115–130.

van Inwagen, P. 1975. "The Incompatibility of Free Will and Determinism." *Philosophical Studies* 27: 185–99.

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

van Inwagen, P. 2004. "Freedom to Break the Laws." *Midwest Studies in Philosophy* 28: 334–350.

van Inwagen, P. 2008a. "How to think about the problem of free will." *Journal of Ethics* 12: 327–341.

van Inwagen, P. 2008b. "The Consequence Argument." In van Inwagen, P.

and Zimmerman, D. (eds.). *Metaphysics. The Big Questions* 2nd. Oxford: Blackwell: 450–456.

van Inwagen, P. 2017. *Thinking about Free Will*. Cambridge, New York: Cambridge University Press.

Vihvelin, K. 2013. *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. New York: Oxford University Press.

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