# Is There Hope for Compatibilism?

John Perry

> …those who accept that responsibility for a situation implies an ability to bring it about and, perhaps, an ability to prevent it, must explain how agents are *able to do* other than they are caused to do. Without it, they can give no defense of their counterexamples. With it, they can be confident that the Consequence Argument, by itself, is no refutation of their position.
>
> Tomis Kapitan (2002, p154)

## 1. Introduction

Compatibilism is the thesis that an act may be both free and determined by previous events and the laws of nature. I assume that in normal cases a condition of a person's performing an act freely is that the person is able to refrain from performing the act. Thus, I accept that if determinism entails that agents do not have this ability, we must give up compatibilism. In this paper I try to contribute to the rethinking of compatibilism by distinguishing between strong and weak accounts of laws and strong and weak accounts of ability. I argue that compatibilism is a tenable position when combined with either a weak account of laws, or a weak account of ability, or both. I shall concentrate on influential arguments for incompatibilism due to Peter van Inwagen, often called collectively the "consequence argument".

Some versions of the consequence argument seem to rely only on inescapable modal principles. In his excellent review and discussion of these arguments, Tomis Kapitan concludes that these principles are not so logically inescapable as to completely trap the compatibilist (Kapitan, 2002). This is not to say van Inwagen's arguments are fallacious, simply that they rely on certain principles about causation and ability that have not yet been fully articulated and defended. Kapitan says, just before making the remarks quoted above, that his assessment provides the compatibilist with "momentary breathing room at best" (2002 p. 154). I am trying to take advantage of the momentary breathing room afforded by Kapitan to explore and to a certain extent defend options

available to the compatibilist. Using terms I explain below, my position is that van Inwagen's arguments do show that the combination of compatibilism with a strong account of laws and a strong account ability (as I define these terms) is not tenable. The options, then, are a weak account of laws, a weak account of ability, or both.

## 2. Basic Argument

As a preliminary to considering a simple argument against compatibilism, let's look at an even simpler argument.

1. $\Box t\ \Box x\ (\Box(x,t) \rightarrow \Box(x, t+1))$

2. $\Box$(Elwood, $t$)

3. Elwood eats cookies at $t+1 \rightarrow \sim\Box$(Elwood, $t+1$)

4. So, ~(Elwood eats cookies at $t$ +1)

Here $\Box$ is a complex state a person together with a suitably large surrounding region the person can be in at a time. If Elwood is in $\Box$ at $t$, then at $t$ states of Elwood and the things around him make it the case that he *really really* does not want a cookie at $t$, that no one is about to persuade him to change his mind, that there are no forces about to impinge on him that will change his mind about cookies, or force him to eat one whether he wants to or not, and so forth. For simplicity we suppose that time is discrete, and $t$+1 is the next instant after $t$. $\Box$ is a state of Elwood and a suitably large region that will succeed $\Box$ according to the laws of nature. It makes it the case that Elwood does not have any cookies in his mouth.

1-4 is a valid argument. If Elwood is in a state that invariably leads to a state that precludes eating cookies, he will not eat cookies. However, it seems quite clear that we are not warranted in going further. To add to our argument:

5. What's more, Elwood *cannot* eat a cookie at $t$+1

would turn it from a valid to an invalid argument.

Prior to being given some quite persuasive argument, we do not take *doing* something as a necessary condition of *being able* to do it, or *being able* to do something as a sufficient condition for *doing it*. We do not take a person's *not* doing something as a sufficient condition for their being *unable* to do it. We accept the inference from

"cannot" to "does not" and from "does" to "can". But we do not accept the inference from "does not" to "cannot" or from "can" to "does".

If Elwood *did* eat a cookie at *t*+1, that would prove that one of the premises 1, 2, or 3 is false. But the fact that he *can* eat a cookie does not show that one of the premises is false. It merely shows that *if* he did do that thing that he is capable of doing, one of the premises *would* be false. Hence, the truth of the premises rules out his eating a cookie, but not his having the ability to eat a cookie.

Now suppose we add another premise, to the effect that (1) is not merely a true universal generalization, but something that follows from the laws of nature. We derive (1) from premise (0):

(0) According to the laws of nature, $\Box t \Box x (\Box(x,t) \rightarrow \Box(x, t+1))$

Now we have the basic argument underlying those used by many incompatibilists, although there are many variations on the basic theme.

The intuitive idea is that premise (0) provides enough extra strength to the premises to get not only to establish (4), that Elwood *does* not eat a cookie, but to establish (5), that he *cannot* eat a cookie. Van Inwagen often taps our intuitions that we cannot change the past or change the laws of nature. The idea is that if Elwood can eat a cookie, then he can falsify one of the premises. He cannot falsify any true statements about the past, so he cannot do anything at *t*+1 that changes the fact that he was in state $\Box$ at *t*. And he cannot falsify laws. So Elwood not only will not, but cannot eat a cookie at *t*+1, if determinism is true.

Given this way of looking at the argument for incompatibilism, there is one basic strategy for the compatibilist. This is to deny that the replacement of (1) with (0) adds enough strength to the premises to validate the step from (4) to (5). There are two basic (and compatible) ways to implement this strategy: (i) adopt a weak theory of laws, and (ii) adopt a weak theory of ability, of *can*, and hence a strong theory of *cannot*. The compatibilist must maintain that it takes more extra power to rule out Elwood's being able to eat the cookie than supplementing (1) with (0) provides, by adopting one or both of these positions.

## 3. Some Preliminaries

Discussions of compatibilism usually employ, at least implicitly, two different concepts having to do with propositions and truth. Propositions *are* true or false. And

propositions *are made* true and false by actions and other events. It will be helpful to be clear about these concepts, and how they are related, before plunging into the main items of business.

Intuitively, many propositions are *made* true, or *rendered* true, by events that occur at some time, or through some interval of time. For example, the proposition *that Nixon won the 1972 election* was not made true or rendered true until the events of Election Day, 1972, or perhaps not even until inauguration day, 1973. What was the status of this proposition before then?

One intuitive option is that propositions are not true or false until they are *made* true or false by events. On this view, many of our statements about the future express propositions that are neither true nor false when they are made, but become true or false when events make them so in the future. This option, though intuitive, does not mesh easily with the two-valued logic that most of us are taught and find easy and convenient to work with.

If we want our familiar logic to mesh smoothly with the concept of events making propositions true, the simplest way is to think of truth and falsity themselves as timeless properties of propositions, while the properties of being rendered or made true or false are properties occur at times, or through intervals. So we have *two* properties having to do with truth of propositions: being true or false, and being made true or false. The first is not relative to times, the second is. This is the track I shall follow in this essay. All of the points I make, however, could be made in the more complicated system, in which some propositions have no truth value until they are made true by events.

How are the properties of being true and being made true related? The obvious way is that:

- if a proposition is ever *made* true, it *is* true.

It might be better to say it *be* true, using a tenseless form, and in fact I shall do so from now on, and say things like

- If a proposition is ever made true (or made false), then it *be* true (or be false).
- The fact that a proposition has not yet been made true (or made false) by events, does not imply that it *be* not true (or be not false).

Compare:

The fact that Gore has not yet been chosen as our next President does not imply that he is not our next President

If Gore ends up being nominated and elected in 2004, then he is our next President. If I call him "our next President" now, I'm correct if the future goes one way, incorrect if it goes the other. It is possible to become rather puzzled by this. How can Gore be our next president now, if it hasn't been decided yet? So he must not be our next President. By parity of reasoning, no one is our next President. That will be a constitutional crisis. But we can avoid the crisis. Being our next President is a property Gore has if at some point between now (summer, 2002) and January 2005 he is elected and inaugurated and Bush hasn't been replaced in the meantime. It all works out, as long as we are careful about the difference between being our next President, a property Gore may have, and being elected and inaugurated as our next President, one which he does not yet have as I write this, in 2002. The fact that lots of propositions be true that have not yet been made true is sort of like that. It can be confusing. It may well be that from a metaphysical point of view our two-valued logic of propositions may not be optimal. Still, if are careful, things will work out.

It will be useful to have the following locution available:

- Events *establish whether P* if they make *P* true or make it false (make *~P* true).

Perhaps we should simply say "make whether *P*," but that sounds even worse.

It seems like there are lots of important propositions whose truth-value is established not by being made true by events, but in some other way. For example, consider Pythagoras's Theorem, the proposition *that the square of the hypotenuse of a right triangle is equal to the sum of the squares of the other two sides*. No event has ever *made* this proposition true, and none ever will. It's not at all like the proposition *that Nixon won the 1972 election.* There is no sequence of events, ending at a certain time, the occurrence of which makes Pythagoras's Theorem true. So the converse of the principle above isn't right; it is not true that if a proposition is true, then some events either have made, are making, or will make it true.

It would not be correct to say that the truth of the proposition that Pythagoras's Theorem is *independent* of events; events do *conform* to it. But they don't *make* it true. They *reflect* its truth.

For propositions that report ordinary facts, such as the proposition *that Eisenhower was president in 1954*, or *that Gore will be president in 2006*, it is natural to use the term "because":

- The proposition *that Eisenhower was president in June, 1953* be true *because* of events that occurred prior to 1954, including his election in 1952 and inauguration in 1953.
- The proposition *that Gore will be president in 2006* be true, if it be, partly because of events that have yet to occur.

With propositions such as that Pythagoras's Theorem, a quite different kind of explanation of their truth seems appropriate, and of course there is a lot of philosophical controversy about what the correct explanation might be. I'll simply say that such propositions are *not* made true by events, and leave it at that for now.

Finally, and importantly, suppose that a true proposition, that is not made true by events, together with some other propositions, that have already been made true by events, entail a proposition that has yet to be made true. What should we say about that proposition? To return to our example, *suppose* that the laws of nature are not made true by events, and that these laws, together with propositions made true by events that have already happened, entail that Elwood will not eat cookies at future time *t*. I will say that although the proposition that Elwood will not eat the cookie at *t* has not yet been *made* true, it's truth value has been *settled*. The proposition won't be made true until the events that the laws of nature and the past determine have actually occurred: the time *t* has arrived, and Elwood's says "no" to the proffered cookie, keeps his arms at his sides, and walks away. But these events were *already* entailed by a combination of propositions some of which were already made true and the rest of which aren't made true by events at all. So the truth-value of the proposition that he would not eat was, in that sense, *already settled*, before he refrained from eating.

The two issues on which I believe the tenability of compatibilism turns are:

- Is the truth of laws *established* by the events that confirm them and fail to disconfirm them, so that laws are laws because events conform to them? Or is the truth of laws established by something else, so that events *conform* to them because they are laws? The first view is a *weak* theory of laws, the second a *strong* theory of laws.
- Can one have the ability to perform or refrain from an action *A* at time *t*, even though the issue of whether one will perform *A* at *t* or refrain from doing so has

been *settled* before *t*? A weak account of ability will allow us to answer *yes* to this question; a strong account will force us to answer *no*.

## 4. Strong and Weak Laws.

One option for the compatibilist is to adopt a weak conception of laws. On a weak conception of laws, (0) does not add much, if anything, to the argument of section 2. Laws are basically true generalizations, and true generalizations are made true at least in part by the events that, as we say, confirm them. The laws that determine that Elwood won't eat a cookie may be true, but, nevertheless, not be made true until the last human or the last cookie has passed out of existence. Elwood's not eating cookies was part of the sequence that established the law, not something the law settled. The law and the facts leading up to Elwood's choice may have *determined* that Elwood would pass up the cookie, but they did not *settle* it, for the law itself wasn't made true until long after Elwood made his decision, and in fact his decision was part of what made it true.

A person can make true generalizations false in the following sense: the generalization be true, but there is something the person could have done, or can do, such that if they had done it, or were to do it, the generalization would not be true. Suppose there were two soccer teams, Manchester United and Nottingham Sherwood. Suppose that Nottingham Sherwood existed from 1960 to 2000 and then was disbanded. During that time it played Manchester United eighty times and never won or tied. So it is a true empirical generalization that

*G*: For all soccer games $g$, if $g$ is a game between Manchester United and Nottingham Sherwood, Nottingham Sherwood loses $g$.

While by 1975 or so this may have seemed like decree of God to the Nottingham fans, we'll assume for now that it was really just a sad but true generalization. Suppose that in the second game between these two teams in 1978 they were in a 0 to 0 tie at the end of the game, and then Manchester won in overtime. In the last second of regulation play, Nottingham had a clear and easy shot on Manchester's goal, which their best player missed. Nottingham's worst player was watching a plane overhead skywriting advertisements for Guinness, and ran into their best player just as he kicked. I think the Nottingham fans at least would think that the Nottingham team could have won that game, even though they did not. And that means they could have made the true generalization *G* false.

I think it is common sense to suppose that laws are not simply true generalizations. Suppose that one Nottingham fan says "We could have won that game in 1978..." Another particularly bitter Nottingham fan says,

> No, you are wrong. You have not grasped that it is an *unshakable law of nature* that Manchester United beats Nottingham Sherwood. It's a law of nature because God decreed that Nottingham would always lose. Laws of nature are universal generalizations that God issued as fiats during creation week, and other things that follow from them. For reasons finite mortals can't be expected to fathom, he often punishes the virtuous and rewards the wicked. And in this spirit he decreed that Manchester United would always beat Nottingham Sherwood. It seemed like Nottingham could win, but in fact it could not."

This fan's remark embodies the intuition behind the argument of section 2. Intuitively, laws are *more* than true generalizations, and (0) adds something substantial to the argument. You can make a mere generalization false; even if no one gets around to doing so, it remains true that someone could have. But laws are laws. You cannot make a generalization false, if it follows from the laws of nature. This is the strong conception of laws. But of course, one can have a strong conception of laws without believing in an unfathomable God or any god at all.

Let's remind ourselves of (1):

(1) $\Box t \Box x (\Box(x,t) \rightarrow \Box(x, t+1))$

If (1) is true then there have not been, and will not be, times $t$ and individuals $x$ such that $\Box(x,t)$ and $\sim\Box(x,t+1)$. And (0) says that (1) is not simply true, but true according to the laws of nature.

The question is, does (1) make (0) true, or does (0) make (1) true? Is (0) true *because* (1) has no disconfirming instances? Or does (1) not have disconfirming instances *because* of the truth of (0)---because (1) follows from the laws of nature? Is the truth of (0) one of those things, like the truth of Pythagoras's Theorem, that is established by something other than what happens? Is it the sort of things to which events conform, but do not make true? Or is it just a sort of fancy way of saying (1)?

To return to our soccer fans. A third soccer fan, also a fan of Hume's, may say:

> I don't think (G) is a decree of God. But I agree with you that it is a law of nature. A law of nature simply is an exceptionless generalization that we have grown used to so that it shapes our expectations. And (G) certainly is that.

This remark would express a *weak* conception of laws. It isn't *quite* enough for (G) to be a law that it is true. More is required: that we use it to form our expectations. But that's all. There is no big metaphysical condition, like a command from God, which is also required. Being a law is just being a true generalization that we have internalized so that our expectations about the future are shaped by it. On this conception, it is (1) being true that explains, or partly explains, that it is a law. (1) Explains (0), not the other way round.

On either conception of laws, laws will have no disconfirming instances. On the strong conception, the fact that *L* is a law *explains* why events conform to it; the truth of the law is due to something *other than* the lack of disconfirming instances. If determinism is also true, laws and propositions about the past not only entail propositions about the future, but also settle them.

On the weak conception of laws, however, the incompatibilist argument does not work. (0) adds nothing to the argument that might push the conclusion from (4) to (5). A Humean conception might add something to the requirement for being a law. (1) not only has to be true, but accepted and used to guide our expectations. But this doesn't push us from (4) to (5).

One option for the compatibilist, then, is to insist on this very weak, Humean conception of laws. What we do is up to us; laws are merely those descriptions of what we do that will end up being true once human actions are complete. Laws determine, but do not settle. I'll call this view "existentialist compatibilism".

I find existentialist compatibilism very appealing, but not wholly convincing. Consider the law that for every action there is an opposite and equal reaction. On the weak conception, this is a law because there never have been and never will be any exceptions to it, and we are attuned to it: when we see a reaction, we expect an opposite and equal reaction. There is nothing about things that *make* this law true, except that everything conforms to it. It seems to me much more plausible that this law *gets* at something (or some things) about the universe that explains why things conform to the law and it has no disconfirming instances. I find it hard to stick with the Humean conception of laws.

One non-theistic but strong conception of laws holds that they are true generalizations that derive from the nature of things, and so describe constraints that form the structure of the world. These constraints are relations between types of things and types of situations.

We can look on (1) as telling us that a certain relation holds between two types of events: co-instantiation. Whenever there is a $\Phi(x,t)$ type of event, there is a $\Phi(x,t+1)$ type of event. On the constraint view, there are other relations between types, such as *causing*, *making happen*, and *forcing*. These are the relations Hume wanted to reduce to, or eliminate in favor of, co-instantiation plus psychology. He called them "necessary connections". I think that "necessary" is rather confusing, given the uses of the term that are familiar in current philosophy. Causal relations are not necessary in the sense of being analytic, or in the sense of being true in all logically possible worlds, or even all metaphysically possible worlds. Still, causal relations between types of events are basic to the structure of the actual world. So (0) explains (1), not the other way around. I'll call them *structural* connections. Structural connections are not necessary, like Pythagoras's Theorem, but they are not made true by events, either. Events conform to them because they capture factors about the world that shape events, not because they report events.

On this conception, if a generalization is true according to the laws of nature, it reflects a constraint that holds among types of things or situations in virtue of their nature, or a necessary consequence of such constraints. Laws are rooted in the nature of matter, space and time, or the nature of whatever else it is that makes up the universe. When one billiard ball hits another the direction and velocity of the second is determined by the direction and velocity of the first in a certain way. What makes this so? Not some decree, sentence, statement, or proposition that truly describes it. Nor the facts about what has happened in similar situations in the past and will hold in the future. There are real connections between types of things and situations. This seems to be what Hume denied, or at least denied we could ever understand. Disagreeing with Hume makes me nervous, and I find it hard to say what else there is about the universe, other than the flow of events, that constitutes such constraints. Nevertheless, I can't bring myself to accept any *weaker* conception of laws.

On this conception, the states of Elwood that are involved in his being in $\Phi$ cause him to not eat cookies. These states include such things as really really not wanting to eat cookies, and seeing no reason to eat cookies. It seems to me that in this sort of case we are in touch with properties that cause us to take or not take certain actions in a

pretty direct way. It seems in the nature of things that someone in such states would not take a cookie. It would be nicest, from a compatibilist point of view, to have a completely weak conception of laws. Nevertheless, this conception seems to hold some promise of fitting into a compatibilist picture, when combined with a suitably weak conception of ability.

## 5. Ability and action

Consider this principle:

(S) If *x can* perform *A* at *t*, then at no time earlier than *t* is it *settled* whether *x* performs *A* at *t*.

A strong theory of can supports (S), while a weak theory does not. I'll argue for a weak theory, and explain why it undercuts van Inwagen's arguments. First an analogy.

It's 1956 and Elwood *doesn't* buy a new Edsel. He thinks they are ugly, ungainly, and overpriced. He doesn't want one. So he doesn't buy one. Now does it follow that he can't afford one? Of course not. He may have all the money he needs, and simply not want one. One question has to do with what he wants in a car and what he thinks about the Edsel. These facts, what he thinks about Edsels and what he want in the way of a car, are pretty much located in Elwood's head. At any rate, they are not located down the street at the bank. But that's where the facts about how much money he has in his account, and how much credit the bankers will give him, reside. It may be that Elwood would rather be drawn and quartered or have rats gnaw out his eyes than buy an Edsel. But those facts about his mind don't tell us anything about his bank account. He may be loaded, so he can easily afford a fleet of Edsels. He can't buy the car without money or credit, but he can *not* buy the car even though he has plenty of money and plenty of credit.

We could put forward a little argument that Elwood won't buy an Edsel:

(1) Reasonable people don't buy cars that they think are ugly, ungainly and overpriced and that they simply don't want and have no other reasons to buy (law of nature).

(2) Elwood thinks Edsels are all of those things, and has no other reason to buy one (fact about Elwood's mind).

(3) So he won't buy one (fact about Elwood's action).

No conclusions about Elwood's bank account can be validly drawn from this argument. It would be silly to draw the further conclusion

(4) So he can't afford one (fact about Elwood's bank account).

The premises don't say anything about Elwood's bank account or his credit. So no conclusions about his bank account or credit can be validly drawn.

That's the model for a weak account of ability. Whether Elwood performs *A* is one question, having to do with what he wants and believes. What he can do is something else, having to do with what abilities he has. If Elwood can't perform *A* then he won't. But it doesn't follow from the fact that he won't that he can't

Of course, people's basic abilities aren't kept down the street in the bank. Elwood's ability to pay for an Edsel may depend in part on his bank account. But his ability to write a check, or say, "Please, sell me an Edsel," depend on facts about him. Still, facts about abilities are quite different than facts about desires and beliefs. Let us suppose, in order to keep this important point vividly in mind, that one part of the brain has to do with what people actually do, and another has to do with what they can do. Let's say the left side contains the desires and beliefs and the other stuff that actually motivates actions. The right side contains all the basic abilities, the repertoire of actions.[1] The repertoire is tapped when one decides to do something that requires a certain ability.

When I learn how to do something, to walk or pick up a glass of water or ride a bicycle or write a check or balance a checkbook or prove a theorem, things change on the right side of my brain. My repertoire of abilities increases. As I learn to do these things, the left side of my brain may change too. I may develop aversions to proving theorems and balancing check books, while learning to like riding bicycles, walking, writing checks and drinking. Then we can predict that I'll do a lot more riding bikes, walking, and drinking than theorem proving and checkbook balancing. What I want to do, and so what I will intentionally try to spend my time doing, depends on the left side. What I can do, depends on the right side.

Given a weak account of ability, the facts that someone did not do something, in the way that we would describe as "intentionally and of his own free will" if we were not worried about determinism, and that his not doing that thing fell under a strong law of nature linking what one thinks and wants with what one does, could hardly have the implication that he *could* not do it. That he does not do it has to do with the

lack of reasons he has for doing it, a fact about the left side of the brain. That he can do it depends on what is going on in the right side of the brain, a quite different question.

If we put an account of abilities in the context of a theory of intentional action, the weak conception of ability makes a lot of sense. Here is a sketch of what is involved in an intentional action; the sketch is no doubt simple-minded and controversial, but I do not think adding sophistication and resolving disputes should affect the points I make (See Israel, Perry and Tutiya, 1993 and Goldman, 1970).

First, there is a *motivating complex of cognitions.* Such a motivating complex for an action *A* includes beliefs (broadly construed, so beliefs include fleeting perceptual beliefs, implicit beliefs, and so forth) and desires (broadly construed, to include wants, urges, whims, and so forth) that rationalize *A*-ing. A set of cognitions *C rationalizes* an action *A* if *A*-ing will promote the satisfaction of the desires in *C* given the truth of the beliefs in *C*. In other words, when a person does something intentionally, there are a bunch of beliefs, perceptions, wants, desires, preferences and the like, which for convenience, I'm just calling "desires and beliefs", relative to which it is reasonable for him to do it. For example, if I intentionally order a vanilla ice cream cone, the motivating complex might include the desire for a vanilla ice cream cone, the perception of a counter, a server, cones, and vanilla ice cream, knowledge of English, a belief that I can afford it, a belief that it won't do me any harm, a belief that I can get one by ordering it, a belief that I can get one by asking for it, and so on.

The motivating complexes cause volitions to perform basic actions, which will cause the basic action, if it is in the repertoire of actions---that is, if the person has the ability to do it. I think of the basic actions as bodily movements, and so use the term "execute" for the special case of performing one of these basic actions: we execute movements, and thereby do lots of other things. I'll try to order a vanilla ice cream cone, by trying to execute coordinated movements of voice-box, throat, lips, tongue and the like that produce the sounds like that will be recognized as the English sentence, "May I please have a vanilla ice cream cone?" *If* that is one of the things I *can* do, I'll say it.

These basic actions in turn cause various results, depending on the circumstances. And these results cause further results, depending on wider and wider circumstances. My words will cause events in the air between me and the ice-cream server, in his ear, in his brain, and so on, until with a little luck I get my ice cream cone.

Here's another example. I am on an airplane and desire a drink of water, and a steward comes by and holds out a tray full of water glasses. I believe that there is a glass of water on the tray in front of me, due to perception and trust in airlines to fill glasses with water rather than gin or hydrochloric acid when they intend to offer them to passengers as water. I know that in these circumstances taking a glass from the tray and drinking from it is a way of quenching my thirst. I can't think of any reason not to take a drink of water. My beliefs and desires rationalize the action of taking a glass from the tray and drinking it, for this will satisfy my desire for a drink of water, without leading to any untoward consequences, given the truth of my beliefs.

This complex will then cause a volition to move in a certain way. Picking up a glass from a tray is a rather delicate movement, but even a klutz like myself can usually do it right. I can suit the movement to the situation based on perception, so that my hand moves to the glass and then brings it to my lips. An important piece of evidence that I can do this is that when I intend to get a glass of water, and see the glass in such and such a relative position, I usually move my hand in a way that succeeds in grabbing it and getting it to my lips. This is due to the ability to execute various movements, and know-how on my part that allows me to execute the right movements in the right situation based on perception. This is something I've gotten reasonably good at, due to years of picking up some things and dropping others.

If there is a glass of water there, and it is reachable in the ordinary ways, and I have the ability to execute the needed movements in the circumstances, then I *can* take a drink of water. If there is a invisible shield between me and the glass, or if the steward is a smart aleck who will move the tray when I get close to it, or if he is an evil airline demon who has brought around glasses full of hydrochloric acid, then I cannot get a drink of water. So, part of the question of whether I can do it is a matter of the circumstances I am in. The other part has to do with what actions are in my repertoire. If I cannot reach as far as I need to, or grab the glass as firmly as required, then I cannot get a drink.

A person has to ability to *bring it about that R in circumstance K* if i) the person's repertoire of basic actions includes some movement *M* such that ii) executing *M* in *K* will have the result that *R*. These conditions for being able to bring it about that *R* can be met when a person does not in fact bring it about that *R* or even try to. Neither of the conditions depends on the person actually bringing it about that *A*. Neither of them require that he *want* to do so, or *have a reason* to do so. They do not preclude the person

*really really wanting not to A.* The person may be willing to die rather than perform *A.* Conditions i) and ii) clearly can be satisfied even if the person's not executing *M* falls under a law of nature to the effect that a person with his motivating complexes will not execute *M.*

This weak account of ability does not support (S). On this account of ability, it does *not* follow from the fact that the (strong) laws of nature plus Elwood's beliefs and desires *settle* that he will not raise his hand at *t*, that he does not have the *ability* to raise it at *t*. That this does not follow can be seen by considering our argument (0)--(5). With a weak theory of ability, it clearly does not work, even if we assume a strong theory of laws. Go down the steps. From (0) to (4) *nothing* is said about abilities. Then, in step (5), abilities are ruled out. It is a left-brain argument, with an invalid right brain conclusion tacked on. It sounds sort of intuitive, but it just doesn't follow.

## 6. Van Inwagen's arguments

### Changing the past

Now let's turn to one version of van Inwagen's argument (2001, p. 23). The issue at hand is whether or *J* could have raised his hand. *Q* is a proposition that rules out *J* having raised his hand at *t*. I'll just take *Q* to be the proposition that *J* did not raise his hand at *t*. Call the Laws of Nature *L* and call the facts up until *t* that are relevant *PF*. Assume that we cannot change the past. Then,

(1) If determinism is true, then the conjunction of *PF* and *L* entails that *Q*.

(2) If *J* had raised his hand at *t*, then *Q* would be false.

(3) If (2) is true, then if *J* could have raised his hand at *t*, *J* could have rendered *Q* false.

(4) If *J* could have rendered *Q* false, and if the conjunction of *PF* and *L* entails *Q*, then *J* could have rendered the conjunction of *PF* and *L* false.

(5) If *J* could have rendered the conjunction of *PF* and *L* false, then *J* could have rendered *L* false (since *J* cannot change *PF*).

(6) *J* could not have rendered *L* false.

(7) So *J* could not have raised his hand.

This argument, if valid, clearly generalizes to any action whatever, and so rules out compatibilism.

Premise (4) is false. The appearance of truth is due to ambiguity in "renders a proposition false". We have to resolve the ambiguity in favor of the sense of "renders false" in which step (3) follows from (2). In this sense (4) is false.

One sense of "render a proposition false" is to do something, which *changes* a proposition from being true, to being false. This seems quite impossible. Of course one can do something that makes a proposition that had *looked like* it was going to turn out to be true turn out to be false instead. It might seem virtually certain that one team is going to win a baseball game, but then the other team scores twelve runs with two outs in the ninth inning to win 12-11. They snatch victory from the jaws of defeat. But they do not really manage to change a proposition from true to false.[2]

It is crystal clear that in this sense of "render false," (3) does not follow from (2). In this sense, (3) says that if *J* could have raised his hand at *t*, then *J* could have changed the truth-value of *that J does not raise his hand at t* from true to false. But this does not follow from (2).

The second sense of "render a proposition false" is to do something which *makes* the negation of the proposition true, at a point in time at which nothing has yet *made* the proposition true or made it false. This concept of "render a proposition false" makes perfectly good sense. By eating a cookie at *t*, I render the proposition *that I do not eat a cookie at t* false. So, the proposition *that I do not eat a cookie at t* be false, and it be false because at *t* I [will refrain/refrain/did refrain] from eating a cookie.

In this sense of "render a proposition false," (3) does follow from (2), as van Inwagen's argument requires. But (4) is false. It does *not* follow from the fact that one renders a proposition *Q* false, in this sense, and that some other proposition *R* entails *Q*, that one also renders *R* false.

Let R be the following proposition. Recall that *Q* is the proposition *that J does not raise his hand at t.* Let *t* be some day after 1950:

*R*: *Q* & *J*'s mother ate a carrot in 1944.

This proposition entails that *J* does not raise his hand at *t*. *J* can render the proposition *Q* false by raising his hand at *t*. If he renders *Q* false, *R* be false, too. But *R* may have

already been rendered false by the time *J* renders *Q* false. This will be the case if *J's* mother did not eat a carrot in 1944. In this case, *J* will not render *R* false, even though *R* entails *Q* and he renders *Q* false. It simply does not follow from the fact that *J* will render *Q* false that he renders false every proposition that entails *Q*. What does follow is that there is no true proposition that entails *Q*.

Principles (i) and (ii) are clearly true, given the coherent concept of "render true" and "render false".

(i) Suppose one does something that renders *P* true. Then no proposition that entails the falsity of *P* be true.

(ii) Suppose one does something that renders *P* false. Then no proposition that entails *P* be true.

Principles (iii)* and (iv)* do not follow, however:

(iii)* Suppose one *can* do something that would render *P* true. Then no proposition that entails the falsity of *P* be true.

(iv)* Suppose one *can* do something that would render *P* false. Then no proposition that entails *P* be true.

Principles *(iii) and *(iv) simply amount to the principle that there is no difference between being able to do something and doing it---that *can* collapses into *does*, and *does not* into *cannot*.

If I can drink a beer, I can render *that I drink a beer* true. So, given (iii)*, if I can drink a beer, no proposition that entails *that I don't drink a beer* is true. So if I can drink a beer, *that I don't drink a beer* isn't true (since it entails itself), so it's false, so I drink a beer. If I can do it, I do it. Can implies does.

Suppose I don't drink the beer. Then, *that I don't drink the beer* is true. Then something is true that entails the falsity of *that I drink the beer.* Then, by (iv)*, I can't render it true that I drink the beer. So I can't drink the beer. Doesn't implies can't.

Such a collapse of "can" into "does" and "doesn't" into "can't" is, of course, just what the incompatibilist wants and the compatibilist needs to avoid. If we accept (iii)* and (iv)*, the collapse would be completed *without any appeal to determinism at all*. But of course there is no reason to accept (iii)* and (iv)*. On the contrary, it seems quite clear that on the weak conception of ability, (iii) and (iv) are true instead:

(iii) Suppose one *can* do something that would render *P* true. This does *not* imply that no proposition that entails the falsity of *P* is true.

(iv) Suppose one *can* do something that would render *P* false. This does *not* imply that no proposition that entails *P* is true.

Suppose *J* can raise his arm at *t*, but decides not to. Then *that J does not raise his arm at t* is true. This proposition entails itself. So *J* can raise his arm at *t*, even though a proposition that entails that he does not raise his arm at *t* is true. So (iii) is correct.

Suppose *J* can refrain from raising his arm at *t*, but in fact he raises it. Then *that J raises his arm at t* is true. This proposition entails itself. So *J* can refrain from raising his arm at *t*, even though a proposition that entails that he does raise his arm at *t* is true. So (iv) is true.

Now let's return to the crucial part of van Inwagen's argument:

(1) If determinism is true, then the conjunction of *PF* and *L* entails that *Q*.

(2) If *J* had raised his hand at *t*, then *Q* would be false.

(3) If (2) is true, then if J could have raised his hand at *t*, *J* could have rendered *Q* false.

(4) If J could have rendered *Q* false, and if the conjunction of *PF* and *L* entails *Q*, then *J* could have rendered the conjunction of *PF* and *L* false.

Since van Inwagen's argument proceeds by very general principles, it should work for any more concrete example we choose. So let:

*PF* = that at *t*-1 *J* really really does not want to raise his hand in the next instant.

*L* = that no one who at *t*-1 really really does not want to raise his hand in the next instant, raises his hand at *t*.

*Q* = that *J* does not raise his hand at *t*.

This example meets the conditions of van Inwagen's argument. That is, *PF* & *L* entails *Q*.

We can certainly accept steps (2) and (3), given our understanding of "render *Q* false". But step (4) does not follow.

*PF* is the proposition that at *t*-1 *J* really really wants to not raise his hand in the next instant. So (4) says that if *J* could render *Q* false (i.e. if he could raise his hand at *t*), then he could render false the proposition:

*that L & at t-1 J really really wants to not raise his hand in the next instant.*

But there is nothing that *J* can do at *t*, the doing of which would *make it the case* that it was not true at *t*-1 that he really really wanted not to raise his arm at *t*.

If *J* does raise his hand at *t*, that will *show*, given *L*, that *PF* is not true. However, that will not *make PF* untrue; it will not *render PF* untrue. If he raises his hand at *t*, that will be because he is in some state at *t-1* than really really wanting not to raise his arm, perhaps in the state of wanting to raise it. In this case, *PF be* untrue, but it be untrue because the events at *t-1* made it false, not because of what *J* does at *t*,

If we go back to our simple picture of what it is for be able to raise your hand at *t*, this should be reasonably clear. There are two facts about *J* and raising his hand, with these possible combinations:

| | **Does raise his hand** | **Does not raise his hand** |
|---|---|---|
| **Can raise his hand** | 1. *Possible* | 2. *Possible* |
| **Cannot raise his hand** | 3. *Not possible* | 4. *Possible* |

The argument starts with the premises that *J does not* raise his hand, i.e., he is in cell 2 or cell 4. It then hypothesizes that he *can* raise his hand, putting him in cell 2. From this it supposed to follow that he changes the past, since the past determines that he will not raise his hand. But it clearly does not follow, for in cell 2 he does not raise his hand, just as the past determines will happen.

I conclude, then, that as long as we have a weak, but realistic and commonsensical concept of ability, we can be determinists and compatibilists, even if we accept a reasonably strong conception of laws.

**Van Inwagen's □ principle.**

Van Inwagen has produced several arguments for incompatibilism. The one I have discussed is the one that seemed most intuitive and convincing to me. Recently more

attention has been paid to an argument from his book *An Essay on Free Will*, (van Inwagen, 1993: 93-104; see also van Inwagen, 2002). The key principles are:

(□) $\Box p \rightarrow Np$

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

where '$Np$' means "$p$ and no one has, or ever had, any choice about whether $p$."

We are thinking of $p$ and $q$ as propositions, and entailment as a relation between propositions. It seems we should accept,

if $p$ entails $q$ then $\Box(p \rightarrow q)$.

Then if we also accept (□) and (□) we'll have to accept the rule,

(□) From $Np$ and *p entails q* deduce $Nq$.

Principle (□) is fatal to compatibilism. If determinism is true, and $p$ is the conjunction of the laws of nature and the facts up until life evolved on earth, and $q$ is any proposition entailed by them describing an act, no one will have any choice whether $q$.

A holder of a weak theory of action will reject (□) and so be spared from (□). To return to our analogy, the premises of the rule of inference (□) tell us nothing about the right side of the brain, while the conclusion does. On a weak theory of action, (□) is not valid.

Recall the criterion for a strong theory of ability:

(S) If *x can* perform *A* at *t*, then at no time earlier than *t* is it *settled* whether *x* performs *A* at *t*.

To be settled at *t* is to follow from some set of propositions, each of which is either established or made true by time *t*. A strong theory of laws says that laws are either established or were made true a long before humans began doing things. So, given a strong theory of laws and (S) and determinism, no one will be able to perform any act *A* at any time. A weak theory of ability denies (S). The weak theory holds that, since the question of whether a person has an ability at a given time need not be affected by his desires and beliefs, and yet it is his desires and beliefs that, together with the laws of nature, determines what he does, the fact that he will or will not do something does not preclude his having the ability to refrain or not refrain from doing it. The weak theorist thinks that a person can have a choice about something, in the sense that they have the

ability to do it or refrain from doing it, even if that thing is determined by laws of nature that are established and facts that are already made true. The weak theorist, then, having rejected (S), need have no qualms about rejecting (□).

## 7. Lewis' Analysis

In "Are We Free to Break the Laws? David Lewis distinguishes between the following claims: "I am able to do something such that, if I did it, a law would be broken" and "I am able to break a law" (Lewis, 2001, 31ff). Suppose the laws of nature and the history of the world up until time *t*-2 entail that I will not take the glass of water at *t*, but I don't. Suppose, as Lewis does, I cannot change the past. There seem two possibilities:

(a) Something happened at *t*-1 that was contrary to the laws of nature, that is, a "divergence miracle".

(b) Everything up to and including *t*-1 was in accord with the laws of nature, but my action was not.

Lewis thinks the fact that I can take the glass of water implies that I am able to do something such that, if I did it, a law would have been broken at some earlier time, but this requires only (a). He does not think I am able to break a law, which would require (b).

I do not think the compatibilist need suppose that if I were to take the drink, any laws of nature would ever need to have been broken. There are auxiliary premises from Lewis's metaphysics and analyses of causation, counterfactuals, and the like that lead him to this defense of compatibilism. But compatibilism by itself does not force us to the divergent miracles defense, and it does not seem to me the most plausible thing to say about cases in which one has the ability to do differently than one does.

If I had taken the drink, freely and voluntarily, then surely my beliefs and preferences would have been different than they actually were. The most likely difference would be that I was thirsty. Assuming determinism, if I had been thirsty when the drink was offered, then something earlier also would have been different; perhaps I wouldn't have taken a drink at the fountain before stepping on the plane, as I did, because the fountain was broken. And that would mean some earlier state of the fountain and its surroundings had been different. And so on. Tracing the changes back to the Big Bang, perhaps it might be a slight difference in the direction in which one particle began its travels through time. Or perhaps it goes back to a deistic god creating

the initial state of the universe a very little bit differently. Or perhaps it just goes back, infinitely. Who knows? It's certainly amazing and weird and in my opinion somewhat depressing that the trail of differences that would have led to my being thirsty rather than not being thirsty should lead back even a couple of thousand years, much less to the beginning of time, or forever. Still, I can't see why either (a) or (b) is required for me to take the glass. Assuming determinism, it follows from the fact that I can accept the drink and don't, that I can do something such that if I did it either the laws of nature or the events up until that time would have been different than they in fact are. It does not follow that if I did what I can do any law would thereby be broken, or any divergence miracle would ever have occurred, or I would have changed the past in any way. I wouldn't have had to change the past, because, according to determinism and the laws of nature, if I had been thirsty, the past would have been different.

## 8. Conclusion

A compatibilist can evade incompatibilist arguments by adopting a weak theory of laws, or a weak theory of ability, or both. My own inclination is in favor of a strong theory of laws and a weak theory of ability.

Although I believe in compatibilism, I am somewhat skeptical about the truth of determinism. I would be happy if it were not true, for I think that determinism is a doctrine that is not very accommodating to important human hopes and aspirations. I don't think the problem is that it rules out freedom, however. I hope I can address these issues in a helpful way on a future occasion. I'm sure I want to, but I'm not at all sure I have the ability to do so.[3]

## References

[Ekstrom, 2001] Ekstrom, Laura Waddell (ed). *Agency and Responsibility: Essays on the Metaphysics of Freedom*. Boulder: Westview Press, 2001.

[Fischer, 1996/2001] Fischer, John Martin. A New Compatibilism. *Philosophical Topics.* 24 (1996): 49-66. Reprinted in [Ekstrom, 2001]: 38--56.

[Fischer, 1994] Fischer, John Martin. *The Metaphysics of Free Will: An Essay on Control.* Cambridge, Mass.: Blackwells, 1994.

[Goldman, 1970] Goldman, Alvin. *A Theory of Human Action*. Englewood Cliffs N.J.: Prentice Hall, 1970.

[Hume,1748] Hume, David. *An Enquiry Concerning Human Understanding.* London, 1748.

[Israel, Perry, Tutiya, 1993] Israel, David, John Perry and Syun Tuitiya. Executions, Motivations and Accomplishments, *The Philosophical Review* October,1993: 515--40.

[Kane, 2002] Kane, Robert, editor. *The Oxford Handbook of Free Will.* New York: Oxford University Press, 2002.

[Kapitan, 2002] Kapitan, Tomis. A Master Argument for Incompatibiism. In [Kane, 2002]: 127-157.

[Lewis, 1981/2001] Lewis, David. Are We Free to Break the Laws? *Theoria*, Vol. XLVII (1981): 113--121. Reprinted in [Ekstrom, 2001]: 30--37.

[Locke, 1694] Locke, John. *Essay on Human Understanding*. 2nd edition. London, 1694.

[van Inwagen, 1975/2001] van Inwagen, Peter. The Incompatibility of Free Will and Determinism. *Philosophical Studies* 27 (1975): 185-1999. Reprinted in {Ekstrom, 2001]: 17--29.

[van Inwagen, 1993] Van Inwagen, Peter. *An Essay on Free Will*, (Oxford: Oxford University Press).

[van Inwagen, 2002] Van Inwagen, Peter. Free Will Remains A Mystery. In [Kane, 2002]:158-177.

---

[1] The right and left sides are chosen completely arbitrarily, simply as a way of easily visualizing the point. This is not an attempt to fit agency into what is known, or thought, or imagined, or claimed, about the differences between the right and left side of the brain.

[2] Note that even if we adopted the more complicated account of the issues discussed in section 3, so that propositions were neither true nor false until events made them so, making a proposition false would not mean changing its truth value from true to false. If we did things this way, we would have to say that when a set of premises entail a

proposition about the future, the truth of the premises requires that the proposition *will be true*, not that it *is* true.

[3] Early versions of this paper were presented at the 2000 Inland Northwest Philosophy Conference and the Philosophy Department Colloquium at the University of Nottingham. I received helpful criticisms and suggestions from members of both audiences. I received helpful comments from Michael Bratman, Joseph Keim Campbell, Eros Corazza and Michael O'Rourke on later versions. Campbell commented on several versions; he saved me from bad mistakes, and suggested helpful repairs. I am very grateful. Much of what I understand about these topics is due to John Martin Fischer, through many conversations and his works, especially Fischer 1994 and Fischer 1996/2001. None of these folks is responsible for the mistakes that remain, although of course if the whole paper is mistaken, and I'm wrong about everything, and both determinism and incompatibilism are true, I'm not either.