METHODE
Analytic Perspectives
www.methodejournal.org

# A dialogue on Free Will

Peter van Inwagen

University of Notre Dame
peter.vaninwagen.1@nd.edu

*Pages:* 212 – 221

**Much of the recent discussion concerning the problem of free will has been centered on the compatibilism/incompatibilism dichotomy. Do you think the central role attributed to this dichotomy is well deserved? And, if so, which of the two alternatives is preferable in your opinion?**

It's hard to answer this question, because it's not clear what "the problem of free will" is. I have proposed that this name be used for the philosophical problem presented by the following dilemma or paradox (or whatever it should be called). In my use, "the problem of free will" is precisely the problem of how to resolve this paradox:

> If the Consequence Argument is sound, then one's having been able to do otherwise than one did on a certain occasion is incompatible with what one did on that occasion's having been causally determined. If the *Mind* Argument is sound, then one's having been able to do otherwise than one did on a certain occasion is incompatible with what one did on that occasion's *not* having been causally determined. A given bad state of affairs is one's fault only if one was able at some point in one's life to have acted in such a way that that state of affairs not obtain. If one was able at some point in one's life to have acted in such a way that a given state of affairs not obtain, then there was an occasion on which one was able to do otherwise than one did on that occasion.


issn: 2481-0498

> If, therefore, any bad state of affairs is or ever has been anyone's fault, either the Consequence Argument is unsound or (inclusive) the *Mind* Argument is unsound. But
>
> (a) Some versions of the Consequence Argument (e.g., the version set out in my answer to Question 2) and some versions of the *Mind* Argument are very convincing, and no one has ever discovered even a plausible candidate for a flaw in either of them,
>
> and
>
> (b) It is false that no bad state of affairs is or ever has been anyone's fault.

Let's call this piece of text "the statement of the Paradox".

If someone tells me that the problem of how to resolve the Paradox is not "the problem of free will," I'll simply make that person a gift of the words 'the problem of free will'. I am interested in the problem the Paradox presents, and not in how the words "the problem of free will" should be used. Nothing that is to be found anywhere in the vast literature on "the problem of free will" is of the slightest interest to me unless it has some bearing on the problem of resolving the Paradox. (And a great deal of it has no bearing on that problem.)

Note that nowhere in the statement of the Paradox does the word 'free' (or do the words 'freely' and 'freedom') occur. *A fortiori* the phrase 'free will' does not occur in the statement. And nowhere in the statement do the words 'responsible' and 'responsibility' occur – and, *a fortiori*, the phrases 'morally responsible' and 'moral responsibility' do not occur in it. Since these words and phrases do not occur in my statement of the Paradox, and since I do not insist on calling the problem of how to resolve the Paradox 'the problem of free will', debates about what "free will" is or about what "the problem of free will" is or about what the relation between "free will" and "moral responsibility" is are of no relevance to the problem of how to resolve the Paradox.

I can, therefore, say only this in response to the first part of your question:

> *If* "the problem of free will" is the problem of how to resolve the Paradox, then *of course* the compatibilism/incompatibilism dichotomy is *one* of the central elements of the problem of free will – for incompatibilism is the conclusion of the Consequence Argument. (It will be demonstrated in the answer to question 2 that incompatibilism can be stated without using the word 'free' or the corresponding adverb and noun.)

Since I find (some versions of) the Consequence Argument to be very convincing (see the answer to question 2), and since there are no real arguments for


issn: 2281-0498

compatibilism at all, I find incompatibilism to be far more plausible than compatibilism.

**In the last three decades the discussions on the so-called "Consequence Argument" have convinced many philosophers that compatibilism is not a viable theoretical option. What is your opinion on that argument?**

I think that the following version of the Consequence Argument is very convincing.

> We begin with three definitions. In each case, the *definiendum* is a pure term of art, a mere abbreviation of its *definiens*.

Say that a proposition $p$ is *untouchable* just in the case that:

> $p$ 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.

Say that, for any time (=instant of time) $t$, a *t-state proposition* is a proposition that gives a complete description of the state of the world at t. (It will be convenient to assume that, for each $t$, there is only one $t$-state proposition – *the* $t$-state proposition. This assumption is not essential to the argument.)

Say that a human being is *multiply able* if there was an occasion on which he or she was trying to decide which of two (or more) incompatible alternative courses of action (e.g. lying and telling the truth) and was at some point in the course of those deliberations able to perform each of them. (If the courses of action in question were lying and telling the truth: at some point in the course of deliberating about whether to lie or to tell the truth, he or she had the ability to lie and had the ability to tell the truth – which does not, of course, imply "had the ability both to lie and to tell the truth.")

The Consequence Argument has the following seven premises:

1. Every necessary truth is untouchable
2. The conjunction of all laws of nature (= &L) is untouchable
3. For every time $t$, if there were as yet no human beings at $t$, the $t$-state proposition is untouchable
4. There is a time $t$ at which there were as yet no human beings
5. If $p$ is untouchable and if the conditional whose antecedent is $p$ and whose consequent is $q$ is untouchable, then $q$ is untouchable.


issn: 2281-0498

6. If the world is deterministic, then, for every true proposition $p$, and every time $t$, the conditional whose antecedent is $\&L$ and whose consequent is (the conditional whose antecedent is the $t$-proposition and whose consequent is $p$) is a necessary truth

7. If any human being is or ever has been multiply able, then some true proposition is not untouchable.

And its conclusion is:

> If the world is deterministic, no human being is or ever has been multiply able.

The demonstration that this argument is logically valid is left as an exercise for the reader. (Textbook quantifier logic suffices.) Notice that, with the sole exception of 'deterministic', none of the traditional vocabulary of "the problem of free will" occurs in this version of the argument. The most controversial of its premises is, I should think, Premise (5) – a version of "Principle $\beta$".

Do not suppose that because I find this version of the Consequence Argument convincing I regard it as a *proof*. It is my firm conviction that no substantive philosophical thesis can be proved. The function of formal presentations of philosophical arguments (like the present version of the Consequence Argument) is to enable philosophers who are in disagreement about the truth-values of the conclusions of those arguments to focus their discussions on individual premises of the arguments and to ensure that no hidden or suppressed premises escape their attention.

**Assuming that libertarianism as such is a viable position, which of the possible libertarian views (such as those centered on agent causation, indeterminist causation or no causation at all) are preferable?**

I myself have consistently avoided the word 'libertarianism'. If one must use the word, it should be defined in the following way (this definition employs the terms of art that were introduced in the presentation of the Consequence Argument in the answer to question 2; essentially equivalent definitions in other terms are of course possible):

> Libertarianism is conjunction of the two following theses:
>
> - At least some human beings are or have been multiply able
> - If the world is deterministic, no human being is or ever has been multiply able.

The *Mind* argument implies that if anyone is ever multiply able, then there must be more to "the springs of action" than the "mere indeterminist causation" theory of action (and *a fortiori* the "no causation at all" theory) can account for.


issn: 2281-0498

For example, the *Mind* argument implies that if it is undermined whether an agent who is deliberating about whether to lie or to tell the truth will lie or will tell the truth, then it is false that that agent is able to lie and false that that agent is able to tell the truth. (Of course, the agent *will* either lie or tell the truth, but that does not imply that the agent is *able* to do either of these things. If, blindfolded, I must draw a card from a well-shuffled standard deck, I *shall* draw either a red card or a black card, but I have neither the ability to draw a red card nor the ability to draw a black card. And when I *have* drawn – say –, a red card, that result does not imply that I had the ability to draw a red card.)

Might it be that the theory of agent-causation "supplies what is missing" in the mere indeterminist-causation theory of action? Might it be that the flaw in the *Mind* argument is that it overlooks the possibility that some undetermined human actions have been produced by agent-causation?

The answer to both these questions seems to me to be No. For suppose that its now being undetermined whether an agent (who is deliberating about whether to do A or do B) will do A or do B is incompatible both with the agent's being able to do A and the agent's being able to do B. Suppose further that agent-causation will indeed play a role in the agent's action (doing A or doing B, as the case may). The occurrence of such an episode of agent-causation can do nothing to alter the incompatibility of its being undetermined whether the agent will do A or B with both the agent's being able to do A and the agent's being able to do B: if $p$ is incompatible with $q$, then the conjunction of $p$ and any other proposition is incompatible with $q$. But this argument is rather abstract. Let us consider a less abstract argument.

> Suppose that agent-causation exists, that agent-causation is a real feature or component of human action; suppose that an agent is deliberating about whether to lie or to tell the truth; if it is undetermined whether that agent will lie or tell the truth, then it is undetermined whether that agent will agent-cause "lie" or agent-cause "tell the truth"; if it is undetermined whether the agent will agent-cause "lie" or agent-cause "tell the truth," then it is false that the agent is able to agent-cause "lie" and false that the agent is able to agent-cause "tell the truth"; if it is false that the agent is able to agent-cause "lie" and false that the agent is able to agent-cause "tell the truth," then it is false that the agent is able to lie and false that the agent is able to tell the truth.

While I accept both indeterminism and the thesis that human beings are sometimes multiply able, I have no theory of action or ability that purports to explain the compatibility of indeterminism and multiple ability. Other philosophers who think that indeterminism and multiple ability are compatible have proposed theories that purport to explain this compatibility. In my view, their theories – when


issn: 2281-0498

they are intelligible at all –, fail to explain what they are supposed to explain.

**During the last years, a growing number of philosophers and scientists have advocated sceptical, eliminativist, pessimistic, or illusionistic views on free will. What do you think of these kinds of views?**

A lot depends on what these people mean by "free will." In the case of the scientists – well, I've never known a scientist who denied the existence of "free will" to make it clear what he or she meant by the words. Some of the philosophers to whom you allude have presented powerful arguments for the non-existence of what I've been calling multiple ability – arguments that consist in a conjunction of the Consequence Argument (or some similar argument) with the *Mind* Argument (or some similar argument). What philosophers who deny the existence of multiple ability must do is to explain how it could be that – if they are right –, any bad thing that has ever happened has been anyone's fault. If some among these philosophers grasp the nettle with their fists and contend that what went on in the Death Camps and in the holds of slavers plying the Middle Passage was not anyone's fault, I congratulate them on their heroic consistency. If others of them affirm both

> Some bad things that have happened *were* someone's fault

and

> Multiple ability does not exist,

they must explain how to deal with the arguments of the following sort:

> Suppose that Ted was driving drunk and has struck and killed a pedestrian. Alice contends, plausibly enough, that the pedestrian's death was Ted's fault. But Alice's plausible contention commits her to the thesis that Ted *should* have done something other than what he did do – he should not have got drunk, perhaps, or he should not have got behind the wheel of a car while in that condition. And to accept the thesis that Ted should have done something other than what he did do commits one to accepting the thesis that he was able to do something other than what he did do.

**A very recent debate concerns the nature of our pre-philosophical views regarding free will. However, some surveys seem to suggest that we naturally tend towards compatibilism, others that we naturally tend towards incompatibilism. What do you think is the value of this kind of "experimental philosophy" in regard to the issue of free will?**

My view of this matter is pretty much the standard view of those who are suspicious of the surveys to which you have alluded. The value of those surveys depends on how the questions they contain are framed, how those surveyed have


issn: 2281-0498

been "primed," and the order in which the questions are asked – a consideration that is borne out by the inconsistent results of the surveys.

I would also point out that a lot depends on what other beliefs those surveyed have and the logical consistency of those beliefs. Suppose, for example that (i) people do tend to give "compatibilistic" responses to a wide variety of imaginary scenarios, but (ii) also have various, well, *general* or abstract beliefs that, taken together, logically imply incompatibilism. We may suppose that people are generally unaware of any inconsistency or logical tension in their positions, and would be perfectly comfortable with simultaneously explicitly affirming both compatiblism and these general beliefs. (That hardly seems implausible. After all, if you told them some stories, and asked them to classify each story as "possible" or "impossible," I at least would be astonished to discover that a very high proportion of them would classify the following story as impossible: 'In the village of Sacramenia in Spain, there lives an adult male barber who shaves all and only those adult males living in Sacramenia who do not shave themselves'.) What should we say about (i) in the light of (ii)? What *I'd* say is that if (ii) is true, the truth of (i) is of very little philosophical importance.

Let me give a simple example that pertains not to people's beliefs about "free will" but rather to their logical or semantical beliefs.

If you present people "out of context" with a conditional that has an obviously false antecedent and an unrelated and *absurdly* false consequent, and ask them whether it's a true statement, they generally say it isn't. (For example: 'If Hillary Clinton is a Republican, then Los Angeles is the capital of Sweden'.) Or so I've sometimes been told. Let's assume it's true. But suppose that, in addition to asking them about the truth-values of particular sentences, you ask them to evaluate certain very general logical principles. This one, for example:

> Suppose that at least one of two statements – call them A and B –, is true.
>
> It follows that if A isn't true, then B is.

I haven't carried out any empirical research, but I bet you can get most of them to assent to this principle. (If the question confuses them, "prime" them with prior questions about examples like this one:

> I have two brothers, Jack and Jim. At least one of the two lives in California.
>
> So if Jack doesn't live in California, Jim does.)

Making allowances for the informality of our statement of the principle, one can use it to deduce 'If Hillary Clinton is a Republican, then Los Angeles is the capital of Sweden' from 'Hillary Clinton isn't a Republican'. If they indeed assent


issn: 2281-0498

to the general principle, what is the philosophical significance of the fact that they tend to say that 'If Hillary Clinton is a Republican, then Los Angeles is the capital of Sweden' isn't true?

Very little, I should think.

Finally, let us suppose for the sake of argument that it is a well-established empirical fact that almost all people without philosophical training tacitly accept compatibilism (or tacitly accept incompatibilism). What philosophical significance should philosophers interested in the problem, "Which of the two is true – compatibilism or incompatibilism?" ascribe to this well-established empirical fact? An interesting question. Here is a parallel question that, in my view, suggests a parallel answer:

> It is a well-established empirical fact that almost all people without training in probability and statistics tacitly accept the thesis that a slot machine (or fruit machine or pokie) that has not paid out a jackpot in a long time is more likely to pay out a jackpot on the next few pulls than is one that has recently paid out a jackpot. What significance should statisticians interested in the question, "*Is* a slot machine that has not paid out a jackpot in a long time (etc.)?" ascribe to this well-established empirical fact?

**What do you think the relationship is between free will and moral responsibility? With regard to this, do you think that the famous Frankfurt scenarios are crucial for assessing the issue?**

I'd have to translate the first part of this question into my own terms before I could answer it. My translation would be something along these lines: What do you think the relation between multiple ability and the assignment of moral fault or blame for bad states of affairs is? If the first part of the question is so understood, my answer to the second part is: None at all. Frankfurt examples are irrelevant to questions about the apportionment of moral fault or blame for the evils of the world. I'll use an example to illustrate the considerations that have led me to adopt this unpopular position.

Suppose that Frieda has promised to feed my cat while I'm away from home and that she has not done this and that, in consequence, the poor cat has died of starvation. Is it Frieda's fault that my cat is dead? I'd say that it is only if Frieda was able to feed the cat. Now imagine someone's objecting to this position as follows.

> But suppose there were some powerful offstage manipulator *who would* have prevented Frieda from feeding the cat if she had shown any inclination to do so – and who would have interacted with Frieda only in that circumstance. It follows that Frieda was unable to feed the cat. (We may suppose that, if it had not been for the offstage manipulator,


issn: 2281-0498

> Frieda would have been able to feed the cat – although, of course, she would not have exercised or acted on this ability.) Since Frieda never in fact made any move towards feeding the cat, the offstage manipulator in fact did nothing. If these things are true, it's Frieda's fault your cat is dead – and this despite the fact that she was unable to feed it.

I see no merit whatever in this objection. It simply *isn't* Frieda's fault that my cat is dead. It isn't her fault for this simple reason: the cat would have died of starvation during my absence no matter what she had done during that period. ("It wouldn't have died if she had fed it." Yes, and it wouldn't have died if she had miraculously conferred on it the ability to survive without food. Don't make the phrase 'no matter what she had done' useless.) One might reasonably conclude from the story that Frieda is an awful person, but the question wasn't what kind of person she was; the question was whether it was her fault that the cat is dead.

**Given the evidence coming from neuroscience and genetics, during the last few years a growing number of scholars have been arguing that the idea that we deserve blame for our bad deeds (and punishment for the worst of them) is ungrounded and should be abandoned. What is your opinion of this view?**

I'm not sure I understand this thesis:

> The idea that we deserve blame for our bad deeds. . . is ungrounded and should be abandoned.

So I'll address instead this thesis, which I do understand:

> The idea that any bad thing that has ever happened was anyone's fault is ungrounded and should be abandoned.

I don't know whether a growing number of scholars – or any scholars at all –, have been arguing in favor of that thesis. If any have, I have to ask whether they would accept its obvious logical consequences, such as

> The idea that the death of six million Jews in the Camps was anyone's fault is ungrounded and should be abandoned.
>
> The idea that the vast amount of human misery caused by the institution of slavery in the American South was anyone's fault is ungrounded and should be abandoned.
>
> In 1979, George Alan was convicted of his wife's murder and sentenced to life in prison. Sixteen years later, he was exonerated and released from prison when it came to light that the police and the district attorney, who had been eager to secure a conviction for a notorious murder, had conspired to conceal evidence that contradicted the theory of the


issn: 2281-0498

crime that the district attorney intended to present to the jury. (In 1997, DNA evidence conclusively proved that the murder had been committed by a man who was then in prison in another state.) The idea that the sixteen years George Alan spent in prison were anyone's fault is ungrounded and should be abandoned.


issn: 2281-0498