

Introduction

Van Inwagen on Free Will

I can remember very clearly the first time van Inwagen encountered the problem of free will. In the autumn of 1965 he was talking with a fellow graduate student at the University of Rochester, one Myles Brand, and made some remark that presupposed the incompatibility of free will and determinism. Brand told him – second-year graduate student to first-year graduate student – that most philosophers believed that free will and determinism were compatible, and outlined some of the currently popular arguments for that position. As Athena from the head of Zeus, the argument that van Inwagen was to publish ten years later in “The Incompatibility of Free Will and Determinism”¹ sprang from his head pretty much full-grown – although it made its entrance into the world by way of his ever-active mouth and not by Athena’s rather more unorthodox route.

The argument had its roots in the following reflections. If free will and determinism coexist, then someone is able to do something not contained in that one possible future that is consistent both with the past and the laws of nature. Suppose that Alice, an inhabitant of a deterministic world, is able to do something she is not in fact going to do; suppose, to be specific, that although she is going to remain a prisoner, she is able to escape from her prison. Her ability to escape can be looked upon as an ability to divert the river of coming events into a channel through which it is not in fact going to flow; to realize, that is, or to cause to be actual, a possible future that is not the future that lies before her, to cause to be actual one

This introductory essay is a revision of my “Van Inwagen on Free Will,” a history of my thinking about free will that was published in 2004 (in *Freedom and Determinism*, ed. J. K. Campbell, M. O’Rourke, and D. Shier [Cambridge, MA: MIT Press, 2004], pp. 213–230). The revisions bring the history up to date.

¹ *Philosophical Studies* 27 (1975): 185–199. Reprinted in *Free Will*, ed. Gary Watson (Oxford University Press, 1983); in *Free Will*, ed. Derk Pereboom (Indianapolis, IN: Hackett, 1997); and in *Agency and Responsibility: Essays on the Metaphysics of Freedom*, ed. Laura Waddell Ekstrom (Boulder, CO: Westview Press, 2000).

of those possible futures in which she escapes. And what would these possible but non-actual futures be like? Let's say that the past *of* a possible but non-actual future is that possible past that would be the actual past if that future were the actual future. Any given future in which Alice escapes must either be a future whose past is the actual past and which is discontinuous with that past, or a future that is continuous with its past but whose past is not the actual past (a past different from the actual past all the way back to the Big Bang), or, finally, a future in which the world is governed by laws of nature that are not the actual laws. But Alice can bring about futures of none of these sorts. If it is insufficiently evident as it stands that she can bring about none of these futures, here is an argument. Let 'Clio' be a proper name for the actual past – thus, when I imagine, as I am about to, Alice using the name 'Clio' in another possible world, I imagine her referring to the past as it is in our world, not the past as it is in hers. Similarly, let 'Nomos' be a proper name for the actual laws of nature (which, remember, we are assuming to be deterministic). It would seem that if Alice is able to escape, she must be able in some sense to cause the actuality of or bring about or realize a future in which she could say, and in so saying speak truly, one of the following three things:

There has been a causal break; the present state of affairs is not continuous with the past

The present is continuous with the past, but that past is not Clio; it is some other past (a past different from Clio all the way back to the Big Bang)

The laws of nature are not Nomos, but some other set of laws.

And, obviously, Alice is not able to get herself into a future in which she can say any of these things and be right.

When van Inwagen got round to writing down the argument that had occurred to him in his conversation with Brand – he first did this in a doctoral thesis he wrote under the supervision of Richard Taylor (*de jure*) and Keith Lehrer (*de facto*) – the argument he wrote down did not *look* much like the argument I have just set out. (It looked very much like the argument he would later publish in "The Incompatibility of Free Will and Determinism.") Nevertheless, the central idea of both arguments was the same, and they no doubt stand or fall together.

When van Inwagen had got his first academic job and was trying to publish this argument, he found it extraordinarily hard to do so. The reason was simple: the conclusion of the argument was known to be false. The

Introduction: Van Inwagen on Free Will

3

unanimous position of the referees whose reports were enclosed with the rejection letters could have been expressed in the following words, which I take from an essay by Donald Davidson, words as well known as they are ungrammatical:

I shall not be directly concerned with such arguments [*sc.* arguments for the incompatibility of free will and determinism], since I know of none that is more than superficially plausible. Hobbes, Locke, Hume, Moore, Schlick, Ayer, Stevenson, and a host of others have done what can be done, or ought ever to have been needed, to remove the confusions that can make determinism seem to oppose freedom.²

It's not that no one was an incompatibilist in those days, of course. The same volume that contained the essay from which my quotation from Davidson is taken also contained David Wiggins's "Toward a Responsible Libertarianism." Roderick M. Chisholm was an incompatibilist, as were Richard Taylor, Carl Ginet, Elizabeth Anscombe, and Peter Geach. Still, it can hardly be denied that incompatibilists were thin on the ground in the sixties and early seventies.

Eventually, however, van Inwagen was able to publish two papers in which he argued for the incompatibility of free will and determinism. (He has always suspected that "The Incompatibility of Free Will and Determinism," which had been rejected by many journals, was accepted by *Philosophical Studies* only because Sellars chose Keith Lehrer as its referee, but he has never known whether this suspicion was correct.) His other paper on the subject, "A Formal Approach to the Problem of Free Will and Determinism,"³ had, in respect of publication, the advantage of making use of the then new and exciting apparatus of possible worlds. (This paper was strongly influenced by Montague's "Deterministic Theories";⁴ the impetus for writing it was a suggestion of Rolf Eberle's.)

Van Inwagen summed up his thought on free will on his 1983 book *An Essay on Free Will*,⁵ and has pretty much avoided learning anything about the problem since – other than by sitting about and thinking it over. (The publication of this book by Oxford University Press was due to the good offices of Tony Kenny and Derek Parfit, neither of whom could have had any sympathy whatever with its content, and van Inwagen has always

² "Freedom to Act," in Ted Honderich (ed.), *Essays on Freedom of Action* (London: Routledge & Kegan Paul, 1973), pp. 137–156. The quoted passage occurs in the opening paragraph of the essay.

³ *Theoria* 40 (1974): Part 1, 9–22.

⁴ Richmond H. Thomason (ed.), *Formal Philosophy: Selected Papers of Richard Montague* (New Haven, CT: Yale University Press, 1974), pp. 303–359.

⁵ Oxford: Clarendon Press, 1983.

been grateful to them for their generosity and has tried to imitate it. He has done his best to learn from them the hard lesson that a philosophical book that he regards as thoroughly wrong-headed can nevertheless be a good book.) Van Inwagen likes to think that this book bears a significant share of the responsibility for the fact that incompatibilists are now much more common than they were thirty or forty years ago. In a paper that he read at a conference in the early nineties, van Inwagen made a remark to the effect that compatibilism was the standard view among philosophers. Michael Slote, who was in the audience, said that he thought that, on the contrary, incompatibilism had become the standard view, or at least the majority view. A few years later, van Inwagen asked Ted Warfield whether he thought that was right. Warfield, who comes as close as is humanly possible to knowing what every analytical philosopher thinks about anything, replied that he thought that the majority of analytical philosophers who had actually worked on the free-will problem were incompatibilists, and that the majority of analytical philosophers (full stop) were compatibilists.⁶

There was one passage in *An Essay on Free Will* that, after the publication of the book, van Inwagen became more and more worried about

⁶ According to the PhilPapers Survey (conducted online in late 2009; see David Bourget and David J. Chalmers, “What Do Philosophers Believe?,” *Philosophical Studies* 170 [2014]: 465–500), the philosophers who responded to the survey responded in the following proportions with respect to four options regarding the problem of free will and determinism:

Compatibilism: Accept (34.8 %), Lean toward (24.3 %)

Libertarianism: Accept (7.7 %), Lean toward (6.0 %)

No free will: Lean toward (6.6 %), Accept (5.7 %)

Other: Agnostic/undecided (4.1 %), The question is too unclear to answer (2.8%)

I would make the following points: (i) By 2009, philosophers no longer used the terms ‘free will’, ‘compatibilism’ and ‘libertarianism’ consistently. For example, by 2009, many philosophers had come to use ‘compatibilism’ as a name for the thesis that *moral responsibility* is compatible with determinism, and had come to use ‘free will’ for the thesis that human agents have access to “whatever kinds of alternative possibilities are required for moral responsibility.” In the thesis advanced in the text I meant this by ‘free will’: agents have free will if they are sometimes able to act otherwise than they do, and by ‘compatibilism’ I mean the thesis that free will in that sense is compatible with determinism. A great many philosophers who would now call themselves ‘compatibilists’ do not reject what I call ‘incompatibilism’; these philosophers would say that the question whether free will in the “able to act otherwise sense” is compatible with determinism is no longer thought to be an interesting question, owing to the fact (they suppose it to be a fact) that we now know that free will in *that* sense is not required for moral responsibility. (ii) There is no way to know what proportion of the respondents to the PhilPapers Survey “actually work on” (i.e., devote a significant proportion of their published work to) the problem of free will. This would, of course, be a rather small proportion of philosophers, and, very likely, a rather small proportion of the respondents. Even if all the respondents who said that they accepted or leaned toward compatibilism meant ‘compatibilism’ in *my* sense (and I suspect that a large proportion of them did not), I do not find the result “accept or lean towards compatibilism: 59.1%” surprising or in tension with the thesis I advanced in the text.

as time passed.⁷ (His worries were aggravated by pointed questions from Alex Rosenberg, and also, curiously enough, by a science-fiction novel by Larry Niven called *Protector* – a novel about which van Inwagen is ready to tell you considerably more than you want to know.) Van Inwagen had said in that passage that although no one was able to render a *physical* law false, it seemed that if there were *psychological* laws, and if we had free will, we had to be able to render *these* laws false. But that raises the question: If one has it within one's power to render some proposition false, in what sense can that proposition be a law? As a sort of schematic example of a psychological law, van Inwagen proposed, 'No one who has received moral training of type A in early childhood ever spreads lying rumors about his professional colleagues in order to advance his career'. He imagined someone asking, "Why does the pattern of behavior described in this statement occur if people don't *have* to conform to it?" He answered this question as follows: 'Perhaps it is just the people who have received moral training of type A in early childhood who *see the point* in not spreading lying rumors about their colleagues'.

He gradually came to see, or to think he saw, that this response to the difficulty was facile, and that the difficulty he had his finger on was broader and deeper than the original puzzle about psychological laws. He gradually came to the conclusion that if one was faced with the necessity of doing either A or B, and that if one saw every reason to do A and no reason whatever to do B, then one would simply not be able to do B. From this conclusion it was no great leap to the slightly stronger conclusion that, if one was faced with a choice between A and B, and one was aware of considerations that could be brought in support of both alternatives, and if the considerations that supported A seemed to one clearly and decisively to outweigh the considerations that supported B, then one would simply not be able to do B. Van Inwagen defended, in "When Is the Will Free?,"⁸ the thesis that the general principles about ability that lead philosophers to incompatibilism should lead anyone who accepts them to accept these conclusions as well. And he went on to argue that, since occasions that call for serious deliberation – occasions, that is, on which one is choosing between alternatives and it does not seem to one that (once all the purely factual questions have been settled) that the reasons that favor either alternative are clearly the stronger – at best only a small proportion of the occasions on which we make a choice are occasions on which we make a *free* choice. ("At best" because there may be no free will at all; perhaps determinism is

⁷ See pp. 63–64.

⁸ Chapter 5 of the present volume.

true, or perhaps – as Broad believed – free action is incompatible both with determinism and indeterminism and is therefore a self-contradictory idea.) Van Inwagen concluded that no action is free unless it is the outcome of deliberation in which one considers reasons that support that act, reasons that support various alternative acts, and in the course of which one finds no obvious answer to the question, “Which set of reasons should prevail?” To take one example among many different sorts of possible example of the consequences of this position, if you answer the telephone “automatically,” if you answer the telephone without so much as considering the question whether you should answer it, your act is not a free act: you could not have done otherwise than answer the telephone; you were not *able* to let it ring till it fell silent; it was *not within your power* not to raise the receiver.

After presenting arguments for this thesis, van Inwagen went on to attempt to show that it does not, or does not obviously, have a certain untoward consequence that it might be thought to have. He contended that from the premise that at best a very small proportion of our acts are free acts, the conclusion does not follow that (at best) only a very small proportion of our ascriptions of moral responsibility are correct. For, he maintained, although there is an inseverable connection between free will and moral responsibility, this connection, inseverable though it be, can be stretched exceeding fine. An example will illustrate his point. Suppose a man is driving drunk and that a pedestrian suddenly looms before him. He attempts to swerve, but too late: he hits and kills the pedestrian because his reflexes are impaired by alcohol. Compare his case with the case of a sober, able, and alert driver whose car strikes and kills a pedestrian in circumstances in which swerving in time to avoid the pedestrian would have required a reaction time smaller than that allowed by the speed of propagation of human neural impulses. In neither case was the driver able to avoid hitting the pedestrian who suddenly loomed before him, but when we consider the former case, we hold the driver morally responsible for the pedestrian’s death, and when we consider the latter case we do not.

The relevant difference, of course, is that the man whose reflexes were impaired by drink, was, so to speak, *able* to avoid being unable to avoid hitting the pedestrian, and the sober and alert driver was *unable* to avoid being unable to avoid hitting the pedestrian. At the moment he first saw the pedestrian, the drunk driver was unable to avoid hitting him, but he had earlier been able, or so we should suppose if we were judge or jury, to avoid driving with the impaired reflexes that were the cause of his fatal inability at the time of the accident. Van Inwagen suggested that this sort of case could serve as a model for the relation between ability and responsibility.

Introduction: Van Inwagen on Free Will

7

Here is a second case, a case in which the inseverable connection between ability and responsibility, though it remains unsevered, as inseverable connections do, is stretched considerably further than it is in most philosophical examples concerning moral responsibility and the ability to do otherwise. Consider a man who is, in middle age, a corrupt politician and is, owing to his corrupted nature, unable to refuse bribes when he believes there is no significant likelihood of the bribery coming to light. That is how he is, but how did he get that way? Suppose the answer is this: As a young man, he made a certain series of free choices, choices preceded by genuine deliberation, which collectively had the effect of establishing him in settled and unbreakable habits of venality. Van Inwagen argued – guided, I suppose, by Aristotle – that this politician can properly be held morally responsible for the baleful effects on the public welfare of the informal services he renders to his political cronies in return for money. And this despite the fact that he is unable, in middle age, to reject the bribes he is offered. He can properly be held responsible for, say, the deaths of the four children in the fire in the building that wasn't up to code, because he could, as a young man, have avoided becoming the sort of man who would be unable to resist the bribe offered by the slumlord who owned the building.

Several philosophers have disputed van Inwagen's conclusion that the principles that lead philosophers to incompatibilism entail that free acts, if they exist at all, are extremely rare, but van Inwagen has never been able to see any force in their arguments. Although he has published answers to them,⁹ he is of the opinion that no answers were needed; that his original arguments were untouched by the arguments of his critics. One philosopher, who generally disagrees with van Inwagen about free will, Dan Dennett, agrees with van Inwagen that these arguments are unanswerable. As Dan put it, referring to "When Is the Will Free?," "Thank you, Peter, for the lovely *reductio* of incompatibilism."

Now van Inwagen's arguments for this conclusion, whether they are good or bad, presuppose that there is an inseverable connection between moral responsibility and the power to do otherwise, however flexible this connection may be. The inseverable connection is this: If one is morally responsible for anything, it follows logically that one has had a free choice about something. But Harry Frankfurt has presented a famous argument that some have taken to refute this thesis.¹⁰ A significant proportion of

⁹ "When the Will Is Not Free," *Philosophical Studies* 75 (1994): 95–113.

¹⁰ See his classic essay "The Principle of Alternate Possibilities," *Journal of Philosophy* 66 (1969): 829–839.

van Inwagen's work on free will has been devoted to Frankfurt's argument. And Frankfurt's argument is important. If it is indeed true that one might be morally responsible for various things, despite one's *never* having been able to do otherwise than one has done, then the problem of free will loses much of its interest – for the simple reason that most people would find the thesis that we lack free will much less unappealing if this thesis could be shown not to entail that we can never be held to moral account for anything.

I have said that Frankfurt's argument has been taken by some to show that it is possible for one to be morally responsible for something even though one has never been able to do otherwise. The actual conclusion of Frankfurt's argument, however, is this: The so-called Principle of Alternative Possibilities is false, or at least not a necessary truth. (I'll call it the Principle of *Alternative* Possibilities. Frankfurt has recently presented an ill-advised defense of his use of the adjective 'alternate' in his name for the principle. It is, I concede, uncharitable of me to mention this. I'll attempt to atone for my lapse by very charitably saying nothing further about it.) This is the Principle of Alternative Possibilities:

A person is morally responsible for what he has done only if he could have done otherwise.

Van Inwagen has always thought that Frankfurt's argument – which, of course, consists in the presentation of a certain sort of counterexample to the Principle of Alternative Possibilities – has a great deal of force and has never been shown conclusively to be mistaken. His position has never been that Frankfurt's proposed counterexamples to the Principle of Alternative Possibilities fail; his position has been rather that even if these counterexamples succeed, even if the Principle of Alternative Possibilities is false, the existence of moral responsibility nonetheless requires the existence of free will.

I have used the qualified phrases "has a great deal of force" and "has never been shown conclusively to be mistaken" because, although van Inwagen is inclined to think that Frankfurt's counterexamples show that the Principle of Alternative Possibilities is false if it is meaningful at all, he's also inclined to think that it's meaningless. I'll briefly say something about why he's inclined to think it's meaningless. If the Principle of Alternative Possibilities is meaningful, the following must be one of the particular statements the general principle endorses:

Bill is morally responsible for lying under oath only if he could have done otherwise.

Introduction: Van Inwagen on Free Will

9

Van Inwagen has never been able to convince himself that he understands sentences like ‘Bill is morally responsible for lying under oath’. It has always seemed reasonably plain to him that what one is morally responsible for is not one’s acts but the *consequences* of one’s acts, or, more exactly, *certain* of the consequences of one’s acts – for no one would suppose that one could be responsible for *all* the consequences of one’s acts. (When I say ‘it has always seemed plain to him’, I mean that it has seemed plain to him since he first encountered the idea that what a person is morally responsible for is the consequences of his acts and not the acts themselves.)¹¹ And the consequences of one’s acts, it would seem, are members of the same ontological category or categories as the consequences of anything that takes place within the causal and temporal order; whatever ontological category one thinks the consequences of a person’s acts should be assigned to, one should assign them to the same category or categories as the consequences of an earthquake or a scientific discovery or a rise in the prime lending rate. There would seem to be two serious candidates for this categorial office: “concrete event” (for example, Caesar’s death), on the one hand, and something proposition-like on the other, “fact,” perhaps, or “state of affairs” (for example, the fact that Caesar disregarded the soothsayer’s warning, or Caesar’s having chosen to believe that Brutus’ strong republican sentiments would never overcome his friendship with and personal loyalty to Caesar).

Now if what one is responsible for is certain of the consequences of one’s acts (and if, for example, one’s telling a lie and the fact that one has lied do not count as “consequences” of themselves), then it is doubtful whether ‘Bill is responsible for lying under oath’ makes sense, and it is therefore doubtful whether the Principle of Alternative Possibilities makes sense, doubtful whether the sentence that formulates it means anything. Van Inwagen, as I have said, has long doubted whether the Principle of Alternative Possibilities does make sense. But he has also been fairly sure that if he’s wrong about this and the sentence ‘A person is morally responsible for what he has done only if he could have done otherwise’ does express some proposition, the proposition it expresses is false, and that Frankfurt’s counterexamples show this.

To recapitulate, van Inwagen thinks that (a) the Principle of Alternative Possibilities is either nonsensical or false, and that (b) moral responsibility

¹¹ He first encountered it when he read P. H. Nowell-Smith’s essay “Action and Responsibility,” in Myles Brand and Douglas Walton (eds.), *Action Theory* (Dordrecht: Reidel, 1976), pp. 311–322. (See especially p. 315.)

nevertheless requires free will – that if anyone is morally responsible for anything, there must be something that person had a free choice about.

He has defended the latter thesis by presenting other principles than the Principle of Alternative Possibilities that have the consequence that moral responsibility requires free will, and presenting arguments designed to show that these other principles seem to be true and cannot be refuted by counterexamples in the style of Frankfurt's counterexamples to the Principle of Alternative Possibilities.¹² One of these principles is

A person is morally responsible for a certain state of affairs only if that state of affairs obtains and he could have prevented it from obtaining.

(Or this was van Inwagen's original formulation of one of these principles. He later came to think that the principle needed to be revised. The revised principle would look something like this:

A person is morally responsible for a certain state of affairs only if that state of affairs obtains and there was a time at which he could so have acted that that state of affairs not obtain.)

The states of affairs quantified over in this principle are "proposition-like": the state of affairs "Caesar's having been murdered" obtains because certain conspirators stabbed Caesar to death in Rome in 44 BC; but it, that very same state of affairs, could have obtained because Cleopatra poisoned him in Alexandria in 48 – just as the proposition that Caesar was murdered is true because certain conspirators stabbed Caesar to death in Rome in 44 BC and could have been true because Cleopatra poisoned him in Alexandria in 48. This state of affairs may thus be contrasted with the concrete event *the murder of Caesar*, which would not have occurred if Cleopatra had poisoned Caesar in Alexandria in 48, although, in that event, there would have been a concrete event, which does not in fact exist, that would have been denoted by the words 'the murder of Caesar'. (He has also endorsed a principle about concrete events that corresponds to or parallels the above principle about states of affairs.)

Van Inwagen has never seen any need to rethink the position he took concerning this principle when he first formulated it in the late seventies (with this minor qualification: as I have said, he has come to prefer a revised version of the principle), to wit that it is extremely plausible, that it entails that moral responsibility requires free will, and that it cannot be refuted by any adaptation of the counterexamples Frankfurt brought against the Principle of Alternative Possibilities. This last point has been disputed, but van

¹² See Chapters 1 and 6 of the present volume.