1
Free Will: The Basics

1.1 Introduction

Are we morally responsible for what we do? Ordinarily, we take it for granted that we are. You might feel guilty about forgetting your mother's birthday, or declining a friend's party invitation when you knew they really wanted you to go. If your mother forgets *your* birthday, or your friend declines your invitation, you might resent them for it, and you might (or might not) forgive them later; whereas, if a friend does something kind for you, you might be grateful to them. These emotional responses are central to our lives, and they are part and parcel of our conception of ourselves and others as bearers of moral responsibility. If we were to think that we are not really morally responsible for what we and others do, it would be inappropriate to feel guilty when we harm others, or resentful or grateful when other people do things that harm or benefit us.

So, what are the conditions under which someone is truly morally responsible for what they do? That's a big question, and this book focuses on just one aspect of it, namely *freedom of the will*. Our starting point is that one requirement that must be satisfied if someone is to be morally responsible for a given act, such as doing a favour for a friend, is that one perform the act *freely*.

It's easy to see why acting freely looks like a plausible requirement on moral responsibility. Imagine, for example, that it turns out that the reason your friend declined your invitation was that she was coerced into doing it: some deranged enemy of yours, hell-bent on ensuring that your party is a failure, had made it clear that if she were to accept, there would be terrible repercussions for her and all her family. In that case, it would certainly be inappropriate for you to resent her for declining

the invitation, and it would be inappropriate because she didn't decline freely. We might put this in other words by saying that she didn't really have a *choice* about whether to decline.

Another kind of situation that seems to remove or diminish moral responsibility concerns one's mental state at the time of the action. Consider 'crimes of passion', where the perpetrator is so overwhelmed by their emotional state – anger, say – that they lose their ability to control their actions. A woman who has been subjected to domestic abuse over a long period of time finally cracks and – without premeditation – stabs her husband to death. Or perhaps she discovers him in the throes of passion with another woman, and – again without premeditation – kills them both. Such cases are legal and moral minefields, but insofar as we take the perpetrators of such acts to lack or have a lower degree of responsibility for their action, we do so because we judge that they were not fully in *control* of what they were doing. And, again, we might put this by saying that they were not acting *freely*, or *of their own free will*.

A third kind of case concerns the personal history of the person whose action we're evaluating. Imagine that your car is stolen. You very much resent the perpetrator of this crime – let's call her Carly. When Carly is apprehended, however, you discover a lot of facts about her life history. She was born into a family of car thieves, was taught to break into cars at a very young age and was expected to assist her parents in their nightly car-thieving rounds, and was generally lacking in any kind of moral education: she was praised for doing some things (stealing cars, for example) and punished for others, but was never taught or encouraged to think about the reasons why some acts are apparently praiseworthy and others merit punishment. Having learned all of this, you *might* take the view that Carly is not really morally responsible for stealing your car after all: it is really her parents who are to blame for having raised her in the way they did. Carly is disposed to steal cars without even considering whether doing so is right or wrong, but this aspect of her character is one that she was not responsible for forming: it is a result of her upbringing, and she had no control over how that went.

This third kind of case might start to make you feel a little uncomfortable, because, of course, we are all the product of a certain kind of upbringing. Most of us are not brought up to be car thieves – indeed, unlike Carly, we may have been brought up to think carefully about the difference between right and wrong. But that we have that kind of character is just as much a product of our upbringing as Carly's character is a product of *her* upbringing; and – you might start to worry – we are

therefore no more morally responsible for our behaviour than Carly is. After all, we had no more choice over who our parents are than she did. Moreover, we are at the mercy not only of the upbringing our parents subjected us to, but also the genes that they endowed us with; and it is entirely possible that many aspects of our character are also influenced by our genetic make-up. And, of course, the same point applies to our parents, too. We might be inclined to blame Carly's parents, rather than Carly, for her car-stealing habit, but they, in turn, were the products of their own upbringing and genetic make-up every bit as much as Carly was. And again, the same point applies to our own parents as well.

The question all of this raises is: Where, if anywhere, does the buck stop – and, if it stops anywhere, *why* does it stop just there? Here's one way you might try to answer that question. Thinking about Carly, you might say: Well, whether or not Carly is morally responsible for stealing my car depends on whether her upbringing and so on *determined* her to steal the car, or whether instead she still had a *choice* about whether to steal it or not. Perhaps her upbringing strongly disposed her towards a life of crime, but perhaps, nonetheless, she was able to *do otherwise* than steal the car. After all, presumably earlier that evening she was sitting down thinking about what she was going to do later on, and after thinking about it, she decided that she would go out car-stealing. But – we might suppose – she didn't *have* to make that decision. She could have decided to stay at home instead, or go to the cinema, or do whatever else she was considering doing. In that case, she really *is* morally responsible for stealing your car. If, on the other hand, her upbringing and so on really did *determine* that there was only one decision she could make in the circumstances – *viz*, to go out car-stealing – then she *isn't* morally responsible for it.

This brings us to what is sometimes known as 'the problem of free will' (although, as we shall see, there is more than one problem of free will), namely, the problem of the apparent incompatibility of free will with *determinism*. The basic idea here is that if someone is *determined* to behave in the way that they do – if Carly, for example, is determined to steal your car by her upbringing, current circumstances and other factors – then it would appear that they do not act *freely*, since there is nothing else they *could* have done. In other words, they *could not have done otherwise* than what they actually did (in Carly's case, stealing your car). If being fully determined to behave in a certain way by one's upbringing (or whatever) deprives one of the ability to do otherwise, and if being able to do otherwise is required for acting freely, then being so determined is incompatible with acting freely. And, finally, if acting

freely is required for being morally responsible for that act, then being determined to act is incompatible with moral responsibility.

The argument just rehearsed is known as the 'Consequence Argument', and we'll come back to it in §1.4. Philosophers tend to fall into two broad camps when it comes to the Consequence Argument. Some are in fundamental sympathy with it: they think that determinism and free will just can't be compatible. Such philosophers are known, unsurprisingly, as *incompatibilists*. Other philosophers, on the other hand, think there is something fishy about the Consequence Argument. They think that determinism and free will don't need to be at odds with one another. For obvious reasons, these philosophers are known as *compatibilists*. (Actually, this is oversimplifying the situation. Some incompatibilists are motivated not by the Consequence Argument but by other arguments for the incompatibility of free will and determinism. We'll come back to these other arguments in §1.5.)

The Consequence Argument gets its intuitive pull from the thought that acting freely requires, to use Daniel Dennett's apt phrase, 'elbow room'. If I *freely* decide whether to steal a car or go to my friend's party, surely I *could* have made a different decision: the decision not to steal the car, say, or the decision to go to the party after all. And the claim is that determinism rules such alternative possibilities out; hence, if determinism is true, we never act freely.

We might, however, ask from the outset whether this kind of 'elbow room' really is required. Consider Wally, who finds a wallet lying in the street. He knows that there is a police station around the corner, and, without even considering the option of taking the wallet himself or pocketing the cash it contains, he hands it in to the police station. That's just the kind of person he is: he's the kind of person who doesn't steal other people's money, even when that money has been left carelessly lying around in the street – not ever. (Well, perhaps he would consider taking – or even actually take – the money if he were flat broke and needed the money to buy drugs to save his ailing grandmother, or some such. But let's assume that this is not the situation that Wally is in right now.) In other words, given Wally's character, taking the money simply isn't a possibility that is open to him: it's not something he could do. Now, do we want to say that Wally is praiseworthy for handing in the wallet? Many compatibilist philosophers (myself included) – those who think that free will and determinism are compatible with one another – say 'yes'. No matter if his upstanding character is entirely determined by the moral education he underwent as a child; he was brought up to do the right thing, and here he is, doing the right thing, and for the right

reasons. If, instead, we were to take the view that Wally is only morally responsible – and so only praiseworthy – for having handed in the wallet if he was *not* fully determined by his upbringing, via the formation of his character, to do that, then we would have to say that he is not praiseworthy for handing in the wallet. Poor Wally! If only his parents hadn't done such a good job, and had instead left him with some slight inclination to keep the wallet for himself, he'd have deserved a pat on the back for handing in the wallet. As it is, he's just doing what he was determined to do – no pat on the back for him.

You might be inclined to respond to this case by saying that – despite the fact that I've tried to rule this out in the way I set it up – Wally *could* have taken the money for himself, assuming that he is a normal, thoughtful and reflective person (which he is): he was entirely *capable* of acting out of character on this particular occasion. In that case, here 's an exercise for you. Try and think of something you really *could not* do. (It doesn't have to be something you couldn't do in *any* circumstances – just something you couldn't do in particular, e.g. normal, circumstances.) It might be, for example, throwing a kitten off a motorway bridge, pushing a total stranger in front of a bus, unleashing a torrent of expletives at your granny, ... you get the picture. Now ask yourself whether you are morally responsible for refraining from doing whatever it was you just thought of.

If your answer is 'No, I am not morally responsible for refraining from swearing at my granny' (or whatever), then it looks as though you are strongly committed to the 'could have done otherwise' condition being a requirement for moral responsibility. But this commitment may be hard to sustain. Consider, for example, the rather unfortunate dilemma it creates for anyone who is a parent. Parents typically want their children to grow up to be the kinds of people who don't push people under buses or swear at their grannies. On the other hand, they typically (though admittedly probably not many parents think about this explicitly) also want their children to grow up to be bearers of moral responsibility: people who are the legitimate targets of 'reactive attitudes' such as gratitude and resentment. (Probably they would prefer it if there were a lot more gratitude than resentment.) Given your answer to the above question, that's a lot to ask. In line with the first desire just described, we have to bring our children up so that they are strongly disposed not to swear at their granny or push people under buses. However, we also have to make sure that there is still *some possibility*, on any given occasion, that they will do the wrong thing – otherwise, we will fail to satisfy our second desire, that they grow up to be morally responsible

agents. That's a tough call: satisfy the first desire, and you run the risk that you've gone too far and deprived your child of moral responsibility. Satisfy the second desire, on the other hand, and you run the serious risk of your child occasionally behaving in unspeakable ways. After all, if you think you really *could* push someone under the bus when the opportunity arises – which, probably, it frequently does – or, at any given moment, call up your granny and give her an earful of verbal abuse, how come you never, ever actually do those things? It doesn't look like you can consistently say both that you *could* do them and that there is no risk whatsoever that you *will*.

Of course, this might not convince you that we can be morally responsible without having the ability to do otherwise. If so, you will still not be convinced that Wally really is morally responsible (and indeed praiseworthy) for handing in the wallet even though he genuinely could not have done otherwise. The question of whether acting freely requires the ability to do otherwise is one that I'll be discussing in a lot more detail later on.

For now, though, let's return to our question about buck-stopping. Earlier I raised the worry that, supposing that we don't hold Carly responsible for stealing your car, Carly's situation and ours are not relevantly different. You might have a different character to Carly's, thanks to a different upbringing; in particular, you were brought up to pay attention whether your actions are right or wrong, and to act accordingly (though, of course, most of us don't *always* succeed in doing so). But you are just as much a product of your upbringing as Carly is a product of hers, and so if Carly is not responsible for her car-stealing *because* she is not responsible for having the character she has, then perhaps you are also not responsible for what you do, whether it is treating your friends well or badly or remembering your mother's birthday – since, like Carly, you are not responsible for having the character *you* have.

The worry here, then, is not so much that our decisions and actions might be determined by our characters and are unfree *for that reason* (this is a worry you will have if you think that Wally deserves no credit for returning the wallet), but that our character itself might be determined by things that we never had any control over – our upbringing and our genetic inheritance, say. We might (and some philosophers do) take the view that the fact that Wally's character rendered him unable to do otherwise *at the time* does not, in itself, entail that he did not act freely; rather, his action is unfree just if he had no control over the formation of that character – the character that in turn determined that he would return the wallet. So, the worry now is that if our *character* is

determined to be the way it is (by our upbringing and so on), then the formation of our character is indeed out of our control. So, determinism threatens free will not because it robs us of 'elbow room' at the time of the relevant decision (Wally could not have done otherwise *at the time*, given his character), but because it entails that the buck *never* stops with us: the chain of determining factors runs right through our own lives, and back through the lives of our parents, and so on. In other words, if determinism is true, *we* cannot be the 'ultimate source' of our actions.

The compatibilist, however, might be inclined to run the argument in the other direction. Rather than saying that you are no more morally responsible for what you do than is Carly, with her unfortunate upbringing, we might start out by saying that manifestly you *are* morally responsible for how you treat your friends or remembering your mother's birthday. Hence Carly is, equally, morally responsible for stealing your car. The circumstances under which people's moral characters develop differ radically. Carly was brought up without any meaningful moral education: she was (we may suppose) discouraged from reflecting on the moral status of car-stealing, or indeed anything else. You (I hope) were not brought up like that: you (let's assume) were brought up to regard the harming of strangers for your own ends as morally unacceptable, and to think carefully about the consequences before making up your mind what to do. It is admittedly a matter of luck – it's out of your control, at least to a large extent – which kind of upbringing you had. Nonetheless, your behaviour stems from the character you actually have, no matter how it was acquired; and, so long as you are not suffering from some psychological compulsion or mental impairment that robs you of the ability to reflect adequately on what you are doing or to consider the consequences of your actions and modify your behaviour accordingly, you and Carly are, equally, morally responsible for that behaviour. That doesn't preclude you from feeling some sympathy for Carly – it is still true, after all, that she is the unlucky victim of an unfortunate upbringing – but that is not the same as thinking that she is not morally responsible, or is less than fully morally responsible, for stealing your car.

One worry about this line of thought is that while it might seem plausible when it comes to Carly, it's unclear whether it's so plausible when it comes to more extreme cases. Imagine, for example, that last week some evil alien neuroscientists rewired your friend Joe's brain while he was asleep. Prior to the rewiring, Joe was a normal, decent person who cared about his friends and would never steal anything unless in extreme circumstances. After the rewiring, Joe cares only about himself

and has no qualms at all about stealing, except insofar as he doesn't want to get caught. Joe really likes the look of your new laptop, and so he steals it from you.

Is Joe blameworthy for what he has done? Intuitively – or at least, according to a lot of people's intuitions – he isn't. After all, it's not *his* fault that the evil alien neuroscientists turned him from someone who would never dream of stealing his friend's laptop into someone who just goes ahead and does it. But this sounds a lot like saying that Joe is off the hook because he is not responsible for the formation of his character. And didn't I just suggest that lack of responsibility for *that* is no impediment to acting freely and hence responsibly, and hence, while we might feel sorry for Carly, who had the misfortune of a terrible upbringing, she and we are equally morally responsible for what we do? So, on what basis do we ascribe moral responsibility to Carly – and ourselves – but withhold it from Joe?

Of course, there are major differences between Joe's case and Carly's, but the question is, why should these differences be relevant to the attribution of moral responsibility? If we want to hold Carly (and ourselves) to account but not Joe, we're going to have to explain why exactly the intervention of the evil alien neuroscientists renders Joe non-responsible for stealing the laptop, and that reason cannot simply be that he had no control over the formation of his character, since that claim (we have assumed) is equally true of us. Some incompatibilists claim that argument like this – sometimes known as 'manipulation arguments' – provide a powerful reason to endorse incompatibilism. And the basic idea is simply that, in the end, the manipulation that Joe suffers at the hands of the evil alien neuroscientists is really no different in relevant respects from the 'manipulation' that ordinary agents who are fully determined by their genes, upbringing and circumstances are subject to. Hence if we and Joe are equally determined to behave as we do, and Joe doesn't act freely, then neither do the rest of us. We'll return to manipulation arguments, and a related argument known as the 'zygote argument', in §1.5 and Chapter 4.

Let's sum up where we've got to. We've seen two kinds of worry that threaten the thought that people who are *determined* (e.g. by their upbringing, genes, current circumstances) to act as they do act freely and hence morally responsibly. One is the thought that such agents lack elbow room: we can't *freely* do what we do unless we could have done otherwise. No pat on the back for Wally, then, since his good character prevents him from being able to keep the money for himself; and, indeed, no praise or blame is due to any of us, ever, if we are determined

to act as we do by facts that were in place before we were even born. The other is the thought that we cannot act freely if we are not the ultimate *source* of our actions, and that our being determined, since before we were born, to act in the way we do entails that we are not the ultimate source of our actions: the buck needs to stop with us if we are to be accountable, and if we are so determined, then the buck does *not* stop with us. Again, then, no praise or blame is due to any of us, ever, if we are determined to act as we do by facts that were in place before we were born. Most incompatibilists take one or other of these worries to undermine compatibilism.

Compatibilists, by contrast, hold that being determined by our characters, upbringing, or whatever to act as we do is no impediment at all to acting freely. Most compatibilists hold that most of us, most of the time, are morally responsible for what we do, whether or not we are determined to do act in that way. Those who are *not* responsible, or not fully responsible, for what they do – whether it is Joe or Carly or the perpetrator of a crime of passion, or an addict or a kleptomaniac or the victim of coercion or whoever – fail to be responsible not because they are *determined* to act as they do, but because they are determined (if they are determined at all) *in a particular way*. The kleptomaniac, for example, may be *compelled* to steal, but compulsion here does not merely amount to being determined; it's a matter of lacking the kind of *control* over one's actions that normal determined agents have. Of course, compatibilists need to explain exactly what kinds of facts about an agent *do* deliver free will, so that they can explain what it is that's lacking in agents who lack free will (see Chapter 2). The same is true, however – at least to some extent – of incompatibilists, or at least those incompatibilists who hold that actual agents really do act freely at least some of the time. The incompatibilist might, for example, try to maintain that the kleptomaniac is *determined* to steal, while normal people, like Carly perhaps, are *not* determined. But, as we'll see, it's not at all clear that they would be entitled to that claim.

So, who's right – the incompatibilist, who holds that Carly and Wally alike behaved freely and morally responsibly only if they could have done otherwise (either at the time of decision or at some prior relevant point in their lives), or the compatibilist, who holds that the issue of the ability to do otherwise (at least in the sense assumed so far) is irrelevant to freedom and responsibility? Are there any options we haven't yet considered?

A further important question is: What exactly *are* the conditions that are required for acting freely? Compatibilists and incompatibilists

disagree about whether or not our not being determined to act as we do is a requirement on acting freely; but even if can resolve that question, it leaves us with a lot of work to do when it comes to giving a positive account of what acting freely requires. As we've just seen, to say, as compatibilists do, that determinism is no bar to acting freely is not yet to say anything about what *is* required. Similarly, no sensible incompatibilist is going to say that our *not* being determined to act as we do is *all* that is required for acting freely. Imagine that Carly has a tiny coin inside her head, so that she'll decide on stealing your car if it lands heads and go to the cinema instead if it lands tails. Would *that* make it the case that – the coin having just landed heads – she decides to steal the car freely? It would seem not – at least if we hang on to the idea that acting freely underpins moral responsibility. For while Carly is now not *determined* to steal your car – after all, the coin *might* have landed tails – she doesn't seem to have any *control* over whether or not she steals it either: it just seems to be a matter of luck. So, she still doesn't seem to be morally responsible for stealing it.

The rest of this book is devoted to trying to figure out how to answer these questions – not an easy task, as we'll see, and that's one reason why the debate about free will continues to rage. The rest of this chapter is devoted to some necessary scene-setting: getting clear on some basics, and fleshing out the arguments for incompatibilism briefly described above in some more detail.

A note on 'freedom of the will'

The problems that are the topic of this book tend to be advertised as problems concerning 'freedom of the will'. But this is a somewhat arcane expression, and it's worth spending a moment explaining it. From the outset, it was an assumption in modern philosophy that the mind is composed of several 'faculties', which were typically thought to include the faculty of reason, the imagination, the faculty of perception, and the will. ('Modern' in the context of 'modern philosophy' doesn't mean 'recent': the 'modern' period started in the sixteenth and seventeenth centuries with philosophers such as Francis Bacon (1561–1626) and René Descartes (1596–1650).) Each of these faculties has its own distinctive role to play in our mental lives: the job of reason is to acquire knowledge, for example, and the role of perception is to acquire information about our surroundings. So, the different faculties correspond to different kinds of mental activity. In the case of the will, the relevant mental activity is mental *action*; as Descartes puts it in the fourth of his *Meditations on First Philosophy*, 'the will ... consists simply in the fact that

when something is put forward for our consideration by the intellect, we are moved to affirm or deny it, or pursue or avoid it' (1641, 101–2).

In ordinary usage, 'the will' has come to be more narrowly associated with desires. Thus, someone's will, in the legal sense of 'last will and testament', is an expression of what they *want* to happen to their possessions after they die; or someone might be described as 'strong-willed' (stubborn or unlikely to change their mind about what they want) or 'weak-willed' (acting contrary to their judgement about what is the best thing to do, and so unable to align what they *want* to do with their judgement about what they *ought* to do).

In philosophy, talk of 'the will' has mostly disappeared; if you read contemporary philosophical texts that claim to be about free will, you will typically find little if any reference to 'the will' or 'willing' (though as we'll see in §2.5, Harry Frankfurt is one exception to this rule). By and large, philosophers these days talk exclusively about freedom of *action*, where action includes both overt, bodily actions, such as switching on a light or picking up a wallet or declining an invitation (a bodily action in that it involves speaking or writing or emailing), and – especially – mental actions such as *deciding* or (perhaps equivalently) *forming an intention* to do something.

1.2 Determinism v. indeterminism

Given the centrality of the question whether *determinism* is compatible with acting freely, we need to get clear on what the thesis of determinism – and its denial, indeterminism – amounts to. So, what *is* determinism? Well, imagine watching a game of snooker on television. If the players are good, you'll often be able accurately to predict what's going to happen when they take the shot: you can tell by the way the player is lining up the shot that he intends to get the black in the corner pocket and have the cue ball bounce back off the cushion and come to rest aligned with a particular red ball for the next shot, or whatever, and – if he's a good player and it's an easy shot – you can therefore predict that that's exactly what will happen. Of course, our predictions aren't always right – snooker would be a *really* boring game if they were. But when we get them wrong, it's reasonable to suppose that that's because there's some further fact about the situation that we didn't know about when we made our prediction. For example, if you make the prediction just after the player hit the cue ball, maybe you didn't know that he hadn't put quite enough power into the shot to do what he wanted to do, or that he hit it at slightly the wrong angle. If

we knew those things (which one can't always discern from watching on TV), then – maybe – we'd be able to predict with total accuracy and certainty what the final position on the table will be when all the balls have come to a standstill.

Imagine that *in principle* we can predict the outcome of the snooker shot with total accuracy and certainty. What is it that we would need to know in order to be able to do this? First of all, we would need to know all the relevant facts about the situation we're looking at: where all the balls are on the table, exactly how hard, and at what angle, the player hits the cue ball, and also a lot of things that we normally simply take for granted – e.g. that the playing surface is felt and not satin and that the balls are regular snooker balls and not made of cheese. But, of course, knowing all that isn't enough; we also need to know *how* things – primarily, snooker balls – behave. We need to know that *if* the player hits the ball with such-and-such force at such-and-such an angle, then the ball will move in such-and-such a direction at such-and-such a speed; that if a ball hits a cushion at a certain speed and angle, it will bounce off at a particular angle with a particular speed. In other words, we need to know some of the *laws of nature* – namely, those laws that are relevant to snooker.

Now, to say that the relevant laws of nature are *deterministic* is to say that for any given starting position or 'initial conditions' – the player hitting the ball with such-and-such force at such-and-such an angle, together with all the other facts about the positions of the balls, what they're made of, and other details – those laws will specify a unique outcome: given a particular starting position, given the laws of nature only one thing *can* happen. For example, the laws will not (again, given a particular starting position) leave it open whether the white ends up aligned with a red or whether it overshoots. By contrast, if the relevant laws are indeterministic, they will *not* specify a unique outcome for a given set of initial conditions. So, in order to be able to predict the outcome of the shot with total accuracy and certainty – never to get it wrong (which of course in practice is pretty much impossible because we are never able to know *everything* about the starting point) – the laws of nature that are relevant to our game of snooker must be deterministic. Or, to put it another way, they must leave nothing to chance. For if the laws were to leave more than one final outcome open, we would have no way of knowing – short of getting in a time machine, zipping a few seconds into the future and taking a look – *which* outcome would result, even if we knew everything there was to know about the initial conditions.

Determinism is the thesis that (i) *all* the laws of nature have this feature – they always specify a unique outcome for a given set of initial conditions – and (ii) everything that happens in the Universe falls under some law of nature or other. In other words, the entire Universe is just like our imagined snooker table: for any given total state of the Universe, the laws of nature specify exactly what the total state of the Universe will be in, say, five seconds' time. For example, if determinism is true, then, if you knew *all* the relevant facts about the wasp that is currently cruising around my study, and you knew all the relevant laws of nature, you'd be able to figure out exactly where the wasp will be in five seconds' time. Of course, wasps are *much* more complicated than snooker balls, and it's a much more difficult – not to mention dangerous – matter trying to find out what state they're in at any given time, and so in practice we're much worse at predicting their movements. But that's just a practical difficulty: if determinism is true, the laws governing the behaviour of the wasp, together with the current state of the wasp and its surroundings, specify a unique outcome. Given the laws, together with current facts about the wasp and its environment, there's only one place the wasp *can* be in five seconds' time.

We, in turn, are much more complicated than wasps; and, in particular, we have a rich inner mental life that wasps lack. In particular, we, unlike wasps, often deliberate about what to do and form intentions as a result of that deliberation. Nonetheless, again, if determinism is true, this merely a practical difficulty when it comes to predicting human behaviour. It's still true that the laws plus all the relevant current facts about me, right now, entail some fact (I don't know what it is) about exactly what I'll be doing, thinking and feeling in five seconds' time.

Indeed, the laws plus current facts specify exactly what the total state of the Universe will be in an hour's time, and a month's time, and in ten billion years' time – including all the facts there will be at that time about me, including facts about what I'll be doing, thinking and feeling. (Well, it's pretty easy to predict that I won't be doing *anything* in ten billion years' time; we can safely assume that the laws plus current facts *don't* leave open the possibility that I'll still exist then.) So, while it's obviously *practically* impossible to predict exactly what I'll be doing in exactly one month's time, again – just as with the wasp – that's only because we lack relevant knowledge.

If determinism is true, then, given the precise state of your brain (or perhaps the precise state of your mind, if you want to resist identifying your state of mind with a particular state of your brain) at this moment, together with the total state of your environment, the laws of nature

uniquely determine what will happen next – and not just what will happen next (e.g. whether you'll decide on a sandwich or some soup for lunch), but what you'll be doing at noon a week on Thursday, the exact moment of your death, whether you will have any grandchildren and whether those grandchildren will ever steal any cars. (It's not hard to see why this might make you worry about freedom of the will, but let's ignore that for now.)

To put it a bit more formally, suppose proposition P_0 is a proposition that specifies the entire state of the Universe at time t_0 (now, say), and suppose L is a proposition that states all the laws of nature. Determinism is the thesis that there is some proposition, P_1, which specifies the entire state of the Universe at some later time, t_1 (noon a week on Thursday, say), such that the conjunction of P_0 and L *entails* P_1. (So if you *knew* P_0 and you knew L – which of course nobody does, and perhaps nobody could do, even in principle – you would be able to *derive* P_1, and you would therefore know exactly what you, and everybody else on the planet, will be doing at noon a week on Thursday.)

So much for determinism. *Indeterminism* is simply the denial of determinism: if indeterminism is true, then it is *not* the case that the precise state of the entire Universe at a given time, together with the laws of nature, specifies a unique outcome for the entire unfolding of the rest of the life of the Universe. How might this work? Well, without getting too far into the deeply puzzling realms of quantum physics, let's consider the concept of radioactive decay. Radioactive atoms, such as strontium-90, have a 'half-life': the period of time such that there is a 50 per cent probability that the atom will decay during that period. (High doses of radiation – that is, a lot of radioactive atoms decaying – are very bad indeed for humans: they cause radiation sickness and longer-term health problems such as leukaemia. And some radioactive atoms have a very long half-life: in the case of strontium-90, it's 28 years. That's the reason why nuclear waste has to be stored safely over a very long period of time, but not indefinitely: after enough time has passed, it's overwhelmingly likely that almost all of the radioactive particles will have decayed. It's also part of the reason why nuclear accidents are to be avoided.)

Now, what does it mean to say that there's a 50 per cent probability of a strontium-90 atom decaying within 28 years? Often, when we ascribe probabilities (or 'likelihoods' or 'chances') to events, our doing so merely reflects our ignorance of further facts. For example, if you have just shuffled a deck of cards, tell me you're going to deal me the top card, and ask me the probability that it's an ace, my answer is 1/13: I know there are 52 cards and four of them are aces. If you had already dealt out the first 20 cards and there were no aces amongst them, my answer would be

4/32 (that is, 1/8). But, of course, in each case, my answer only reflects my *ignorance* of what the top card is. There is a fact of the matter about whether it's an ace; it's just that I have no way of knowing what that fact of the matter is. So, there is also a perfectly good sense in which the probability of the top card being an ace is either 1 (it's an ace) or 0 (it isn't) – I just don't know which. While this is a perfectly good sense of probability, however, you're unlikely to win at cards if you don't adopt the former, ignorance-based, not-1-or-0 probabilities as a basis for your play!

So, is the case of strontium-90 like the case of being dealt an ace? That is, is there some further fact of the matter that determines the precise time at which a given strontium-90 atom will decay? Are seemingly-identical strontium-90 atoms actually different to one another in some way that explains why some decay after 47 days and some after 12,689 days? The safest answer is to say: We don't know. If there *is* some further fact of the matter, physicists certainly haven't yet figured out what it is. Perhaps there is some further fact of the matter (as Einstein, who famously said that 'God does not play dice', believed) – in which case, the laws of nature governing radioactive decay are deterministic, it's just that we don't (yet) know what they are. If there is *no* further fact of the matter, then the laws governing radioactive decay are indeterministic, and so – given that indeterminism is simply the denial that everything is governed by deterministic laws – indeterminism is true, and determinism is false. (Actually, some philosophers – myself included – think that we shouldn't really hold that the laws *govern* what happens at all; see §3.5. But let's ignore that for now.)

There are two really important points to grasp when it comes to thinking about the connection between indeterminism and freedom of the will. The first is that *we don't know* whether or not determinism is true. Perhaps one day the sciences will answer that question, but as things currently stand, we're a long way from knowing. (See Balaguer 2010, Chapter 4 for a good survey.) So – in particular – simply assuming that determinism is true, or indeed that indeterminism is true, is not a good thing to do! However plausible determinism (or indeterminism) might seem to you, whether or not it is true is not, in fact, something you can discern just by considering how plausible it seems to you. Quite a lot of the time in philosophy, we adopt or reject theories on the basis of how plausible they are: by the extent to which they accord with our 'intuitions' or pre-theoretic judgements. I'll come back to the issue of the status of intuitions in philosophical methodology in §7.4, but that issue isn't really relevant here, because determinism is not a *philosophical* thesis at all – it's an empirical thesis. So, the question of whether or not it is true is a question for scientific investigation, not philosophical reflection.

Second, if you think that free will is incompatible with determinism, then the mere fact (if it is a fact) that indeterminism is true does not, just by itself, secure freedom of the will. Grant that, say, radioactive decay is – and perhaps the laws of quantum mechanics more generally are – genuinely indeterministic. That, just by itself, has no implications for whether or not anybody ever acts freely; all it does (if incompatibilism is true) is remove one possible obstacle. But obstacles remain. We'll come back to this point in Chapter 5.

1.3 Determinism, indeterminism and causation

To avoid confusion later on, it's very important to note that determinism is *not* equivalent to the thesis that every event has a cause. Prior to around the middle of the twentieth century, most philosophers – including philosophers writing about free will – *did* take determinism to be the thesis that every event has a cause, presumably because they simply assumed that all causation must be *deterministic* causation. Thus, for example, G.E. Moore says: 'if everything is caused, it must be true, in *some* sense, that we *never could* have done, what we did not do' (1912, 110).

One possible explanation of this conflation is that indeterminism only really started to be taken seriously by scientists in the early twentieth century, when quantum physics began to be developed. It was only at that point that the question of whether events that are not fully determined by the laws of nature plus facts about the past may yet have causes began to be addressed; and these days, most philosophers would agree that undetermined events *can* have causes.

Take our example of radioactive decay. It is surely obviously true that the nuclear accident at Chernobyl in 1986 *caused* people to suffer (and indeed die) from acute radiation sickness in the same year: the accident caused a vast quantity of radioactive material to be released, and that in turn – via radioactive decay – caused acute radiation sickness. But, as we've seen, radioactive decay is (so far as we know) indeterministic. So, the release of the radioactive material did not *determine* that anyone would get sick during 1986, since it's entirely consistent with the laws of physics – though admittedly extremely unlikely – that *none*, or very few, of the radioactive particles decayed that year; and if none or very few had decayed, there would have been no radiation sickness.

To take a more humdrum example, suppose that coin-flipping is indeterministic, so that whether the coin lands heads or tails isn't determined by the starting position (e.g. how you flip the coin) plus the laws. If I bet you £10 that the coin will land heads, you accept the bet, the

coin is flipped, and it lands tails, you're £10 better off. Surely your newfound wealth has causes, including my offering you the bet and the flipping of the coin: if I hadn't offered you the bet, or no coin had been flipped, you wouldn't have got the money. So, again, it looks like we can have *indeterministic* causation: you can *cause* something to happen without *determining* it to happen.

As I said, it's important to be aware that much of the older (by which I mean before about 1970) literature on free will assumes the contrary – that all causation is deterministic causation. Here, for example, is an argument of Carl Ginet's (1962) for the incompatibility of free will and determinism. (a) If our decisions were caused, it would be possible in principle to know in advance what one were going to decide, before one decided it. But (b) it is impossible to decide to do something if you already know what it is you're going to decide. (After all, how can you deliberate about whether to go to the shops or stay in and watch TV if you already somehow *know* you're going to decide to go to the shops? What exactly would you be deliberating about?) Conclusion: Decisions cannot be caused. (Thus, Ginet is sometimes described as defending a 'contra-causal' account of free will.)

We can see that premise (a) above presupposes that all causation must be deterministic causation: it is only if our decisions are *deterministically* caused, or, equivalently, *causally determinined*, that we could in principle – by knowing the laws and all the relevant facts about the past – know what we're going to decide. For if our decisions were *indeterministically* caused – say your deliberation still left it open whether you would decide on going to the shops or staying in – then you could *not* in principle know what you were going to decide, even if you knew the laws and all the relevant facts about the past. So, in fact, assuming Ginet is right about premise (b), his argument only really shows that our decisions cannot be *deterministically* caused – which does not, of course, entail that they cannot be caused *at all*. (We'll briefly come back to the question of the relationship between free will and foreknowledge in §7.2.)

1.4 The Consequence Argument

Now that we've got clear on what the thesis of determinism is, we're in a position to be able to get to grips with perhaps the most famous argument for incompatibilism, which I briefly introduced in §1.1. This argument has been around for a very long time in various different forms, but its best-known formulation is that of Peter van Inwagen (1975). Van

Inwagen's formulation is quite technical. We'll get back to the technicalities in §3.1, but for now we'll just stick to the basic idea, which goes something like this:

First of all – and, as we'll see, this is a crucial premise – van Inwagen assumes that doing something freely requires that I *could* have acted differently. So, for example, in order to freely make a cup of tea, it must be the case that, at some time prior to my actually making the tea, I could have refrained from making it.

Now, as we've seen, if determinism is true, then a proposition stating all the laws of nature (call this proposition L), together with a proposition that describes the precise state of the Universe at a given time t_0 – say, 1 p.m. on 2 January 2004 (call this proposition P_0) – jointly entail a proposition that describes the precise state of the Universe at some later time t – 6 p.m. on 26 March 2012, say. Now, as a matter of fact, one thing that happened at t is that I made a cup of tea. Let P be the proposition that I made a cup of tea at t. So – assuming determinism – P is entailed by (L & P_0). Now, *could* I have refrained from making the tea then, even though in fact I did not refrain and did, in fact, make the tea?

Well, P is entailed by – that is, P is a *consequence of* – L & P_0. So, if I could have refrained from making the tea – in van Inwagen's terms, if I could have 'rendered P false' – then it must be the case that I could also have rendered L & P_0 false. Why? Because if you can render some proposition R false, and R is entailed by Q, then you can also render Q false. For example, suppose there are nine people in the room (Q). That entails that there are fewer than ten people in the room (R), so Q entails R. Suppose I can render R false, for example, by entering the room, thereby increasing the number of people in it to ten. Then clearly I can also render Q false – I can also render it false that there are nine people in the room. Of course, this is just *one* case where some proposition Q entails another, R, and where it's true that if I can render R false, I can also render Q false. Van Inwagen is claiming that this is true of *all* propositions Q and R, where Q entails R. The fact that one instance of this general claim is true does not, of course, demonstrate that the general claim is true. If you doubt that the general claim is true, you need to find a case where (i) Q entails R, and (ii) someone can render R false, but they cannot render Q false. Go ahead and try!

So, if we want to know whether I could have rendered P false, we need to know whether I could have rendered L & P_0 false. Given the above principle, and given that L & P_0 entails P, if I could *not* have rendered L & P_0 false, then it must be the case that I could not have rendered P false – otherwise we'd be contradicting the principle we've just assumed

to be true. So, the question is: *Could* I have rendered L & P_0 false? Well, clearly I could not, at any stage later than t_0 (1 p.m. on 2 January 2004), have rendered P_0 false: what's past is past, and you can't do anything about that. I can't now render false the fact that I had toast for breakfast this morning, or the fact that I just typed the word 'breakfast'; I can delete the word, but I can't make it the case that I never typed it in the first place.

It follows that I could *only* have rendered L & P_0 false if I could have rendered L false. But nobody can render the laws of nature false: if proposition is a law of nature (e.g. if it's a law of nature that nothing travels faster than the speed of light), then nothing we do can possibly render that proposition false. To put it more simply, we cannot break the laws of nature. So, in fact, I could not have rendered L false. And, as we saw above, I could not render P_0 false either. So, clearly I could not have rendered their conjunction – L & P_0 – false. But in that case, I could not have rendered P false either. (Remember: If I could have rendered P false, then I could have rendered L & P_0 false. But since I could not have done the latter, I could not have done the former either.) Hence, if determinism is true, I could not have refrained from making a cup of tea at t. And, since being such that I could have done otherwise than make a cup of tea at t is a requirement on my making it freely, it follows that if determinism is true, I did not make the cup of tea freely.

To put things much more straightforwardly, though rather less rigorously: the laws of nature and facts about the past aren't up to me. So, the consequences of the laws of nature and facts about the past can't be up to me either. So, if (as determinism entails) everything I do is a consequence of the laws of nature and facts about the past, nothing I do is really up to me. Hence, acting freely is incompatible with determinism.

Recall that compatibilists think that acting freely and responsibly is compatible with determinism, and incompatibilists don't. Many – but not all – incompatibilists reject the claim that acting freely and responsibly is compatible with determinism because they are convinced by the Consequence Argument. We'll come back to the highly contentious question of whether or not the Consequence Argument really works in Chapter 3; below, however, I turn to a different kind of argument for incompatibilism.

1.5 Sourcehood and manipulation arguments

Whether we're compatibilists or incompatibilists, it's plausible to think that at least *some* kinds of manipulation restrict our freedom. In the

film *The Truman Show*, Truman Burbank thinks he's an ordinary guy in an ordinary town with an ordinary job. What he doesn't know is that he's spent his whole life living on a giant soap opera set, and all of his supposed friends and family and colleagues are actors. Truman is deliberately placed by the director and scriptwriters in situations that will make for good TV. Truman is clearly the victim of a kind of constant manipulation. While his behaviour in any given situation is not generally any more predictable than yours or mine (by people who know us very well), his circumstances are frequently engineered in such a way as to make it likely that he'll behave in the way that the director wants him to behave. In particular, Truman unfortunately develops the desire to get outside the cosy but restricting (literally restricting, in fact, but Truman doesn't know this) town of Seahaven. The director manipulates Truman in two ways: by thwarting his desires (e.g. cancelling flights, making the bus break down) and by affecting his desires themselves through external circumstances (e.g. putting news reports on TV about the dangers of travelling and killing off Truman's 'father').

What should we say about Truman's predicament? Clearly, in one sense his freedom has been restricted. In particular, he is not free to leave the set: this is something he cannot do (although he does – spoiler alert! – eventually manage to escape). But we might still say, at least up to the point where the director has to start (apparently) cancelling flights and so on, that he freely remains – even though his *desire* to remain has itself been manipulated by circumstances. After all, our desires are manipulated by circumstances all the time – it happens every time we watch an advert or walk around the supermarket. But we still want to say that we freely buy the things we buy, and are morally responsible for those purchases.

But what about more extreme cases of manipulation? Remember Joe the laptop thief from §1.1, whose brain has been interfered with by evil alien neurosurgeons so that Joe's previous good and law-abiding nature has been replaced with a concern only for his own short-term material gain in such a way that he is now determined to steal the laptop. Does Joe freely do so? A simple 'manipulation argument' against compatibilism runs like this. Clearly Joe does *not* freely steal the laptop, and is therefore not blameworthy for doing so. But in that case, the compatibilist needs to explain what exactly the difference is between Joe's case and that of a normal deterministic agent. After all, there are plenty of people who are just like post-manipulation Joe but who have not been the victims of alien intervention, and *they*, according to the compatibilist, are blameworthy for what they do.

Here's a different manipulation-style argument: the 'zygote argument'. Imagine that Ernie lives in a fully deterministic universe. Diana, a goddess with extraordinary powers (and foresight), creates a zygote in Mary – which grows up to be Ernie – in just such a way that Ernie is guaranteed, given the facts at the time plus the laws, to steal his friend's laptop in thirty years' time; and Diana knows this. Indeed, she *intends* for Ernie to steal the laptop, and she knows that by creating the zygote in just the right way she can ensure that this happens, so that's what she does. Ernie grows up in the normal way: he is not subject to brainwashing or alien intervention (aside from the creation of his zygote), Diana does not create the zygote in such a way that Ernie comes to have irresistible impulses or kleptomania, and so on. In other words, creation aside, Ernie is exactly like any normal adult human being (albeit not a very nice one), and hence would seem to satisfy whatever conditions on acting freely that compatibilists might care to name (see Chapter 2).

As with Joe, we're supposed to find it intuitively compelling that Ernie does *not* act freely and responsibly in stealing the laptop. And yet, he would appear to be a normal, fully functioning deterministic adult, and so the compatibilist would apparently have no grounds for *denying* that Ernie acts freely. Hence, since Ernie does not act freely, neither does any other normal, fully functioning deterministic adult. Hence compatibilism must be false. Note that while Diana does not directly intervene in Ernie's life (post-conception) in any way, there still a sense in which Ernie is the victim of manipulation. After all, Diana brought him into existence *intending* that he would steal his friend's laptop, and – given determinism – there was nothing, at any stage, that Ernie could do to stop this happening.

It's clear that the compatibilist faces a difficult choice here: accept that Joe and Ernie are morally responsible for what they do, which is (allegedly) implausible, or else try to find some way of distinguishing between Joe's and Ernie's predicaments on the one hand and, on the other, the situation that normal deterministic agents are in, so that Joe and Ernie get off the hook, but the rest of us (or at least most of us, most of the time) do not. There are various moves the compatibilist might try and make at this point. For example, in Joe's case we might try to claim that there is something about the sudden and wholesale alien intervention that is (obviously) completely unlike what happens to normal deterministic agents, and so we just need to add some kind of historical condition – perhaps a condition that specifies that one's psychology develop in the 'normal' way – in order to deal with Joe's case. Unfortunately, even if we manage to come up with such a condition to

deal with Joe's case, we're not going to be able to apply it to Ernie's case. After all, Ernie's psychology *does* develop in the 'normal' way. Ernie's *zygote* certainly didn't come into existence in the normal way, but it's unclear why facts about *that* aspect of Ernie's history should be thought to render him unfree.

We'll consider manipulation arguments in more detail in Chapter 4. For now, it's worth noting that incompatibilists generally take manipulation arguments, such as those described above, to connect with the issue of *sourcehood* raised in §1.1. Remember Wally, who is determined by his good character to hand the wallet in to the police station so that he cannot, now, do otherwise than hand in the wallet. And recall the worry that if determinism is true, his character in turn was determined to be the way it is by factors that were ultimately outside Wally's control. There is no point at which the buck stops with Wally; the buck passes right through Wally and right out the other side. Manipulation arguments describe cases, such as Joe's and Ernie's, where, again, it seems that the buck doesn't stop with the agent. (Perhaps in Joe's case the buck stops with the evil alien neuroscientists, and in Ernie's case it stops with Diana: these are the people who are *ultimately* responsible for Joe and Ernie behaving as they do.) The force of manipulation arguments lies in the thought that, if determinism is true, there isn't *really* any relevant difference between Joe and Ernie on the one hand, and the rest of us on the other. After all, they are (or perhaps just Ernie is) just like us in all *relevant* respects. So, if the buck passes right through Ernie and out the other side, then the same is true of us, since there is nothing in *us* to stop the buck that is not present in Ernie. If there were, then Ernie would not be the normally-functioning deterministic agent that he is stipulated to be.

1.6 Conclusion

This main point of this chapter has been to introduce you to some of the broad questions and themes that will get discussed in a lot more detail later on. In particular, the various ways in which the compatibilist might respond to the Consequence Argument are discussed in Chapter 3, and compatibilist reactions to 'sourcehood' arguments, such as the kinds of manipulation argument just described, are considered in Chapter 4. But these arguments only address the very general question of whether there are good reasons for thinking that acting freely is incompatible with determinism.

The more significant question, I think, is whether we have any good reasons to think that *we* act freely – we real, flesh-and-blood agents – as we go about our daily lives. This is a really important question. After all, if we never (or hardly ever) act freely, then we are never (or hardly ever) entitled to hold people responsible for what they do: nobody is ever praiseworthy or blameworthy, so nobody is ever *deserving* of praise or blame or – for example – gratitude or forgiveness. That might (conceivably) not make much of a *practical* difference to our lives; perhaps, even in the absence of moral responsibility, we would still have good grounds for rewarding people who treat us well and incarcerating criminals. But it would, or so I think, make a significant *moral* difference to our lives. If there is no such thing as moral responsibility, then there would appear to be no *moral* difference between the friend who misses your wedding because she has a pathological fear of flying and just can't make herself get on the plane despite really not wanting to let you down, and the friend who misses it because, well, she knew it was important to you, but she just couldn't be bothered to make the effort.

So, *is* there a moral difference between the cases? To answer that question, we need to have settled a lot more than just the question whether acting freely is compatible with determinism – we need a *theory* of what acting freely consists in. That will be our concern in Chapter 2, where I consider compatibilist theories, and Chapter 5, where I consider incompatibilist theories.

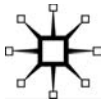
Chapter 6 returns to the question of whether acting freely requires the *ability to do otherwise*, in the sense of my doing otherwise being an alternative possibility that is left open by the past plus the laws. The claim that acting freely *does* require this ability is, as we've seen, a central premise in the Consequence Argument for incompatibilism, but it is thrown into serious doubt by a famous argument of Harry Frankfurt's, and we'll see how incompatibilists have responded to that argument. Finally, in Chapter 7, I briefly consider some additional issues and draw attention to some loose ends.

Free Will

An Introduction

Helen Beebee
Samuel Hall Professor of Philosophy, University of Manchester, UK

palgrave
macmillan



First published 2013 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries

ISBN: 978–0–230–23292–1 (hardback)
ISBN: 978–0–230–23293–8 (paperback)

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

A catalog record for this book is available from the Library of Congress.