Free will and the ability to do otherwise

Simon Kittle

Submitted for the degree of Doctor of Philosophy

Department of Philosophy, University of Sheffield

April 2015

Abstract

This thesis is an investigation into the nature of those abilities that are relevant to free will when the latter is understood as requiring the ability to do otherwise. I assume from the outset the traditional and intuitive picture that being able to do otherwise bestows a significant kind of control on an agent and I ask what kinds of ability are implicated in such control. In chapter 1 I assess the simple conditional analysis of the sense of 'can' relevant to free will, and I agree with the consensus that this analysis fails. In chapter 2 I consider Kadri Vihvelin's contemporary version of the conditional account, which is cast in terms of dispositions. I develop, via engagement with Vihvelin's view, an account of how modal properties such as dispositions and abilities should be individuated. In chapter 3 I use this account to show why Vihvelin's account of free will is unsatisfactory. I also show how it helps us to better understand a distinction that is often made in the free will literature, namely, that between some notion of 'general' ability on the one hand and a notion of 'specific' or 'particular' abilities on the other. I argue that there are two important distinctions, both of which are relevant to free will. In chapter 4 I consider Keith Lehrer's analysis of 'can' and conclude that while it fails to achieve Lehrer's stated aim – namely, that of demonstrating that free will is compatible with determinism – it does contain some useful insights about the kinds of ability relevant to free will. In the fifth and final chapter I switch gears somewhat; I argue that various conditions typically treated under the banner of 'the epistemic criteria on moral responsibility' should instead be treated as conditions on an agent's being able to do otherwise.

Acknowledgements

First and foremost I would like to thank my wife, Hannah, for enabling me to begin, persevere with, and finish this thesis, as well as for the many hours spent listening to and engaging with my ideas.

I owe a huge debt to my supervisors, Eric Olson and Jessica Leech, for many helpful discussions, comments, suggestions, friendly criticisms, as well as much valuable advice concerning this thesis and the pursuit of philosophy more widely. Thanks especially go to Eric, as my primary supervisor, for patiently reading and re-reading, in very careful detail, what is to follow.

I would like to thank Kadri Vihvelin for helpful clarifications on her account of free will (discussed at length in chapters 2 and 3) and Ann Whittle for useful discussions with respect to her position on classifying abilities (discussed in chapter 3). I would also like to thank John Martin Fischer both for helpful discussion with respect to a number of papers related to this thesis and for valuable encouragement in the study of free will more generally.

for Hannah

Contents

Introduction.. 9

Chapter 1 – The simple conditional analysis 15

1.1. The simple conditional analysis of ‘can’ 15

1.2. Problems for the simple conditional account.............................. 23

Chapter 2 – Vihvelin’s dispositional compatibilism............................ 42

2.1. Vihvelin’s account 42

2.2. Vihvelin’s account still plagued by the Transfer Problem 60

2.3. Vihvelin on the nature of abilities 66

Chapter 3 – Abilities and circumstances 75

3.1. A recap 75

3.2. The nature of abilities and how they relate to one another 77

3.3. The kinds of ability relevant to free will................................ 85

3.4. Abilities to do otherwise and Vihvelin’s treatment of Frankfurt-style cases 96

Chapter 4 – Possible worlds based analyses of ‘can’ 101

4.1. Lehrer’s ‘”Can” in theory and practice’..................................101

4.2. Horgan’s donation counter-example 112

4.3. Objection: the notion of an advantage depends on that of ability 117

4.4. Objection: the need for an epistemic criterion on world accessibility 121

4.5. Campbell’s ‘A theory of compatibilist alternatives’ 122

4.6. Campbell’s C4 and the acting freely/freedom to do otherwise distinction ..126

4.7. Campbell’s C4 and causal relevance..128

Chapter 5 – Epistemic abilities..131

5.1. Introduction 131

5.2. The separation of freedom and epistemic conditions on responsibility......132

5.3. ‘Epistemic’ abilities..138

Introduction

There are many things that are outside the control of any human person. And there are things that some people, but not others, have control over. But we ordinarily think that people are at least sometimes – and perhaps most of the time – in control of their own actions. Moreover, the control that people have over their actions is thought to be distinctive. If a computer breaks, we might get frustrated and kick it, but we do not *blame* the computer. Similarly, if a dog tears a sofa to shreds, we might shout at it and banish it from the lounge, but we do not *blame* it. Things are very different when a person wrongs us in some way. The wrongdoer might experience feelings of guilt or regret; and we might feel indignation, hurt, disrespected, resentment and so on. Associated with these attitudes is the idea that the person is to blame. It is not just that the person *caused* something bad to occur; rather, the person was in some more robust sense the *source* or *origin* of the action. The action is therefore significant in the following way: it is a harm done, and the harm is the person's *fault*. Similarly, if someone does something morally good, we feel that the action is attributable to that person in a way which justifies praise. These distinctive responses only make sense if the person possessed a distinctive kind of control. That, at least, is the traditional view, and it is the view to be investigated in this thesis. The kind of control at issue is the one implied when we say that what someone did was *up to them* or that *they had a choice about it*. To have this kind of control is to have free will. But what does this control amount to?

Traditionally, free will has been defined as *the ability to do otherwise*. If a person acts, that person had free will only if they *could have done otherwise*; only if they had *the power to have acted differently*. In other words, a person has free will on a particular occasion only if they are in a situation where there are at least two available options and the person is able to take either of them. This thought has been expressed in various ways: e.g. in terms of what the person *could* do, or what the person *is able* to do, or what it is within the person's *power* to do. Arguably, however, the key idea is the same: free will is about having alternatives and being able to take those alternatives. When someone is able to realise any out of a range of alternatives, what the person does is *up to them*. Thomas Pink describes the idea like this:

> Freedom or control is inherently a power that can be exercised in more than one way – to determine either that a given action occurs or that it does not. We have control over which actions we perform, whereas ordinary causes lack control over which effects they produce (Pink 2004: 118).

On this way of looking at things, when a person has control over some action it is *because* they could have done otherwise or were able to do otherwise. It is possessing this kind of power or ability – the power to settle which action they perform – that makes the person the *source* or *origin* of their actions. This understanding of free will is natural and persuasive. It fits well with the idea that we face an open future, and that by being able to decide this way or that, we are able to settle various things about it. This picture is nicely captured in the following passage from John Martin Fischer:

> We naturally think of the future as open. We think of the future as containing various paths that branch off one past; although we know we will travel along just one of these paths, we take it that some of the other paths are (at least sometimes) genuinely accessible to us. In deliberating and deciding on a course of action, we intuitively think of ourselves (at least sometimes) as determining which path to take, among various paths that we could take. Thus, we think of ourselves as having control over which path we take (Fischer 1994: 190).

In what follows I will be concerned with the best way to make sense of this intuitive picture. Many accounts of what it is to be able to do otherwise are formulated in the context of addressing perceived threats to or problems with the intuitive picture outlined above. Two threats in particular feature prominently in what follows. The first comes from the thesis of causal determinism. Historically, causal determinism has been explicitly articulated in terms of causation: for every event there is a set of prior events or conditions which together with the laws of nature are sufficient for the occurrence of that event. It is more common in recent work to formulate the idea without explicitly mentioning causation. Peter van Inwagen, for example, describes causal determinism as ‘the thesis that the past and the laws of nature together determine a unique future, that only one future is consistent with the past and the laws of nature’ (van Inwagen 1989: 400). Kadri Vihvelin’s characterisation is as follows: the laws of nature specify sufficient (and not just necessary or probabilistic) conditions and those laws govern everything that happens, including all human actions (Vihvelin 2013: 3). Causation is implicated by the mention of the laws of nature, but this more recent way of formulating the doctrine is useful because issues in the philosophy of causation are highly contested. I will accept van Inwagen’s (1989: 400) definition:

> (**Determinism**) The past and the laws of nature together determine a unique future; only one future is consistent with the past and the laws of nature.

This definition is not problem free but however we spell it out causal determinism produces a kind of conditional or relative necessity: given the laws of nature and certain conditions, such-and-such *must* happen. Fischer (1994: 191) is very clear about why this poses a problem. If free will requires the ability to do otherwise, then a person has free will on some occasion only if *it is possible* (in some sense) for that person to do otherwise (on that occasion). But if determinism is true, then only one thing is possible given the laws of the nature and the actual past. So if determinism is true and someone has free will, then that person either has the ability to act such that the past would have been different or the ability to falsify the laws of nature. But this is vastly at odds with our ordinary conception of the world: we ordinarily think of the past as fixed in some strong sense, and we do not typically think that people can break the laws of nature. If we take these thoughts seriously, then if determinism is true we must conclude that no one is ever able to do otherwise and so no one ever has free will. That, at least, is the worry. In a series of writings van Inwagen has famously turned the above worry into a powerful argument for *incompatibilism*: the doctrine that free will and determinism are incompatible. Van Inwagen called it the Consequence Argument; he has presented a number of different versions of the argument but the following informal statement is enough for the current purposes:

> If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born; and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us (van Inwagen 1983: 16).

Many of the accounts of free will to be considered are motivated by a desire to show where this argument goes wrong; that is, they seek to show why determinism is not the threat to free will that it might initially appear to be. Such accounts are therefore *compatibilist*: they hold that free will, understood as involving the ability to do otherwise, is compatible with causal determinism. This thesis engages with four compatibilist accounts of the ability to do otherwise and concludes that all of them fail. I will refer to these accounts as *classical compatibilist* accounts because they accept the classical understanding of free will as involving the ability to do otherwise.

The second threat to this understanding of free will comes from Harry Frankfurt. He presented an argument in 1969 aiming to show that the following principle, which I will refer to as the Principle of Alternative Possibilities (PAP),[1] is false:

> (**PAP**) A person is morally responsible for what he has done only if he could have done otherwise (Frankfurt 1969: 829).

Frankfurt contends, in other words, that the ability to do otherwise is *irrelevant* to moral responsibility. Frankfurt's argument was based on an example with the following form:

> (**A Frankfurt-style case**) Jones is thinking about killing Smith. Black too wants Smith to die, but he'd prefer not to have to do it himself. To that end, he installs a device in Jones's brain which is capable of monitoring and manipulating Jones's brain. Black's device allows him to detect any sign that Jones might show of not deciding to kill Smith. In addition, it allows Black to cause Jones to kill Smith should he detect any sign that Jones is having second thoughts. As it happens, Jones goes ahead and kills Smith on his own.

The idea behind the example is this: Jones, we are meant to agree, is morally responsible for killing Smith. After all, Jones kills Smith 'on his own'; Black was there, ready and waiting, but in the end he didn't need to intervene. However, Black's presence is meant to make it the case that Jones had no options: Jones could not have avoided killing Smith because if he had attempted anything different Black would have manipulated him into killing Smith. Therefore, we are invited to conclude that a person can be morally responsible for an action despite not being able to do otherwise. I will refer to examples of this kind as *Frankfurt-style cases*. Many philosophers,

[1] Frankfurt called it the Principle of **Alternate** Possibilities, but the bulk of the literature following him as called it the Principle of **Alternative** Possibilities.

impressed by Frankfurt's argument, conclude that while moral responsibility does require free will, free will does not require or involve the ability to do otherwise. Fischer (1994) is one such philosopher. He thinks it is 'highly plausible' that the ability to do otherwise is *incompatible* with causal determinism (Fischer 2012: 88). But he also thinks that the control required for moral responsibility is *compatible* with causal determinism because the latter does not require the former. Fischer users the term *semi-compatibilism* to describe this position: the ability to do otherwise is *incompatible* with determinism but moral responsibility – and whatever kind of freedom or control it requires – is *compatible* with determinism. Frankfurt's argument, and Frankfurt-style cases more generally, have been hugely influential, and although I will not address the cases directly, I will discuss them at various points (primarily in 3.4 and 4.5).

One consequence of Frankfurt's argument is that many philosophers, even those who do not in the end think his arguments are successful, now define free will as the control required for moral responsibility *whatever the nature of that control turns out to be*. That is, many now accept that it is far from obvious whether free will involves the ability to do otherwise. Defining free will as the control required for moral responsibility is thought to be a more neutral way of approaching the issues. This approach is right about one thing: there is a close connection between free will and moral responsibility. Indeed, the very reason why it was thought to be obvious, prior to Frankfurt, that free will involved the ability to do otherwise was because it was thought obvious that such was the control required for moral responsibility. For this reason, issues pertaining to moral responsibility will feature at various points throughout.

This thesis is an investigation into the nature of free will on the assumption that it involves the ability to do otherwise. I will primarily be concerned with determining the kind of ability or power properties which are required for free will. This task is complicated by the fact that the terms usually used to refer to the relevant ability or power – 'can', 'able', 'ability', 'power' – either admit of a variety of different meanings or are context sensitive (or both). This is produces difficulties from the start because it means that the target of investigation is not immediately clear. To illustrate: suppose that Olivia is sat in her office and I sneak up and lock the door from the outside. There is a sense in which Olivia is able to leave her office: she's awake, she could decide to leave, she's able to walk, navigate rooms and turn door handles, and so on. Olivia is able to leave her office in a sense in which someone suffering from severe paralysis would not be. But there is also a sense in which she is unable to leave her office due to the locked door. It might seem evident that it is the latter sense which is relevant to Olivia's freedom; after all, it's not up to Olivia whether she leaves and that is precisely because the door is locked. As we will see, however, this thought has been challenged. Furthermore, it's not clear that there are only two uses of 'able' that need to be distinguished. And while it is easy enough to gesture at a distinction using examples such as the one above, providing an account of that distinction is rather more difficult. A large part of my project is to identify the different kinds of ability that a person might possess and then to investigate which ones are relevant to free will.

I begin in chapter 1 by assessing what is known as the *simple conditional analysis* of the sense of 'can' relevant to free will, a view first explicitly articulated by G. E. Moore in 1912. Although I will concur with the wide consensus that the simple conditional analysis fails, the survey of this early debate introduces a number of important issues and sets the stage for the rest of the thesis. In chapter 2 I consider Vihvelin's (2013) contemporary version of the conditional account which is cast in terms of dispositions. I develop, via engagement with Vihvelin's view, an account of how modal properties such as dispositions and abilities should be individuated. Often these properties are individuated using a set of *stimulus conditions* and a *manifestation*. For example, *fragility*, the (philosopher's) paradigm example of a disposition, might be characterised as *the disposition to break when struck*. Here, the *being struck* is the stimulus event while *breaking* is the manifestation event. I argue that, in addition to the event or action type typically cited as the manifestation behaviour (e.g. breaking, walking, deciding), dispositions and abilities should be individuated by, first, a set of circumstances against which that behaviour is to be expected, and second, a 'modal force' parameter which is a measure of the 'strength' of the disposition or ability. In chapter 3 I elaborate on these points by identifying two ways that dispositions and abilities might be thought of as *general*. This is significant because in the free will literature a distinction is often made between some kind of 'general' ability on the one hand and a 'specific', 'particular' or 'local' ability on the other. One of the key claims of chapter 3 is that there are two distinctions – one corresponding to the each of the ways an ability might be general. It is crucial, I go on to suggest, that these two distinctions be kept separate when discussing free will. Employing the two notions of general, I show how there is, for each action type, not just one ability to perform actions of that type, but what we might think of as a spectrum of such abilities. I chart various ways that abilities on this spectrum can relate to each other and show why recognising the spectrum of abilities undermines Vihvelin's compatibilist account of the ability to do otherwise. In chapter 4 I consider Keith Lehrer's (1976) early possibility based analysis of 'can'. I conclude that while it fails to achieve Lehrer's stated aim – namely, that of demonstrating the compatibility of determinism and the ability to do otherwise – it does contain some important insights about the sense in which agents need to be able to do otherwise. In the fifth and final chapter I switch gears somewhat. I argue that various conditions typically treated under the banner of 'the epistemic conditions on moral responsibility' should instead be treated as a condition on the agent's possessing the kind of abilities relevant to free will. These conditions – broadly, those concerning what an agent *believe*s about what they are doing – affect not just what an agent can be morally responsible for but also what it is that an agent can control. In the end, I suggest that an agent is able to do otherwise in the sense relevant to free will only if the agent is has the right *doxastic abilities*: abilities which require certain conditions on what an agent believes to be satisfied.

Chapter 1 – The simple conditional analysis

1.1. The simple conditional analysis of 'can'

1.1.1. Moore and the simple conditional analysis

In this chapter I will assess the *simple conditional analysis* of the sense of 'can' relevant to free will as articulated by G. E. Moore in 1912. Although this view is generally considered inadequate today, it is worth surveying because it introduces a number of important issues. Moreover, the objections raised against this view will have to be addressed by any contemporary view which proposes to explicate the sense of 'can' or 'able' relevant to free will in terms of conditionals. Vihvelin's (2004; 2013) account of free will, discussed in the next chapter, falls into this category. Moore takes the ordinary understanding of free will to be something very similar to the characterisation outlined in the Introduction:

> The statement that we have Free Will is certainly ordinarily understood to imply that we really sometimes have the power of acting differently from the way in which we actually do act (Moore 1912: 203).

He goes on to say that if an agent does not have it within their power to act otherwise then that agent does not have free will (Moore 1912: 203–4). The inverse entailment does not follow: that an agent can do otherwise does not imply that they have free will. Moore says this, not to emphasise that there are other conditions (e.g. epistemic) which an agent must satisfy in order to have free will, but because he takes 'can' to be ambiguous. It might therefore be that an agent can do otherwise or has the power to act differently, but still does not have free will, because the sense in which the agent can (has the power to) do otherwise is not relevant to free will. This highlights an important assumption Moore is making, namely, that there is just one sense of 'can' or 'able' that is always relevant to free will. I will accept this assumption for the purposes of this chapter, and will examine it more closely in chapter 3.

The motivation behind Moore's conditional analysis of 'can' is to reconcile free will with the thesis of determinism. Moore does not cast it in these terms, but talks of reconciling free will with what he calls the *Principle of Causality*, namely, the idea that 'absolutely everything that happens has a cause in what precedes it' (Moore 1912: 208). Moore's understanding of the Principle of Causality, however, may reasonably be taken to imply the thesis of determinism. Consider the following argument which he puts in the mouth of someone who objects to the reconciliation project:

> [Assume] ... that absolutely everything that happens has a *cause* in what precedes it. But to say this is to say that it follows *necessarily* from something that preceded it; or, in other words, that, once the preceding events which are its cause had happened, it was absolutely bound to happen. But to say that it was *bound* to happen, is to say that nothing else *could* have happened instead; so that, if *everything* has a cause, *nothing* ever could have happened except what did happen (Moore 1912: 208–9).

It is clear from this quotation that Moore's understanding of *cause* is such that causes necessitate their effects. Together with the assumption that everything has a cause, determinism follows. And although Moore leaves it unsaid, the worry here is the one outlined in the Introduction: 'if everything has a cause, nothing ever could have happened except what did happen', and this would seem to imply that no one has (or ever had) the ability to do otherwise and so no one has (or ever had) free will.

Moore begins his reconciliation of free will and determinism by stressing that if we grant the premisses, what follows is that in *one sense of the word* 'could' it is true to say that nothing ever could have happened save for what did happen. What doesn't follow is that there is *no other sense* of 'could' in which something else could have happened. Moore's account of free will exploits this point by taking the word 'could' to be ambiguous. He seeks to show that the sense of 'could' related to free will is a sense in which it can be true to say of agents that they could've done otherwise even given the truth of the Principle of Causality. He begins by presenting an example which he uses to construct his analysis. Suppose, Moore says, that he neither went for a walk nor for a run this morning. Next consider two actions: Moore's walking a mile in twenty minutes this morning and his running two miles in five minutes this morning. Moore *did* neither of these things but there is an important difference between them: Moore *could* have walked a mile in twenty minutes this morning but he *could not* have run two miles in five minutes (Moore 1912: 206).

What is the sense in which Moore could have walked a mile in twenty minutes but could not have run two miles in five minutes? Moore suggests that one very obvious answer to this question is that 'could' here means 'could, if I had chosen' (Moore 1912: 211). After all, if Moore had chosen to walk a mile it's likely that he could have completed it in twenty minutes, but the same is not true of running two miles in five minutes. Now, given his use of the word 'could' in the analysans one of two things must be the case: either (a) Moore thinks that there is a further ambiguity in 'could' such that the occurrence in the analysans is a different sense of the word to that in the analysandum (both, of course, being different to the sense in which, given the truth of determinism, the agent could not have done any different), or (b) he thinks that the 'could' statements we are attempting to explain require a complementary 'if' clause *in order to complete the sense of the sentence*. In the latter case, Moore's suggestion above would not be an *analysis* of 'could' statements but merely the first stage in isolating the form of the sentence to be analysed.

Shortly after introducing this idea Moore notes that due to a 'possible complication' it might be better to say that 'could' means '***should*** have, if I had chosen' (Moore 1912: 211). Moore never discusses this 'possible complication' but it's reasonable to assume that the 'complication' he has in mind is the use of 'could' – the term he is attempting to analyse – in the analysans. Later on in the same chapter of *Ethics* he introduces a third option, saying that '[t]here are certainly good reasons for thinking that we very often mean by "could" merely "would, if so and so had chosen"' (Moore 1912: 212). Right from the outset then, we have three options:

(**CA-C**) 'I could have A-ed' is true if and only if I **could** have A-ed, if I had chosen to A
(**CA-S**) 'I could have A-ed' is true if and only if I **should** have A-ed, if I had chosen to A
(**CA-W**) 'I could have A-ed' is true if and only if I **would** have A-ed, if I had chosen to A

The following questions come to mind:

Does Moore intend to read **CA-C** as employing a third sense of 'could' (option (a) above) or as a way of spelling out the incomplete meaning of 'could' statements (option (b))? And are either of these readings plausible?
What, if anything, is the difference between **CA-S** and **CA-W**?
If there is a difference between **CA-S** and **CA-W**, which version is to be preferred?

I will answer these questions in section 1.1.3.

1.1.2. Some varieties of 'could' statements

Before answering the questions listed above it is useful to explore, with the help of J. L. Austin (1979), why we have two options when it comes to interpreting **CA-C**. Austin pointed out two important and orthogonal distinctions which it is useful to discuss, not just for purposes of this chapter, but also for those which follow. The first distinction is that between indicative and subjunctive uses of 'could', the second is that between statements which employ 'could' and those which employ 'could have'.

In the indicative mood, statements and clauses employing 'could' (without the 'have' modifier) are in the past tense and are capable of being complete sentences: they do not require supplementing with an additional clause in order to express a complete proposition. For example, 'I could swim 5km', when read as being in the indicative mood, is very similar in meaning to the statement 'I used to be able to swim 5km'. Such a sentence expresses a complete thought and so is understandable as it is, even if it would be clearer to add an explicit reference to when in the past I had this ability, for example, by saying 'I could swim 5km when I was fifteen'.

This contrasts with the subjunctive use of 'could'. When used in the subjunctive mood the statement 'I could swim 5km' might mean something like 'I could swim 5km *if I trained for 2 months*' or 'I could swim 5km *if the sea weren't so choppy*'. In these cases, what I am able to do depends on some further condition obtaining (having trained, the sea being calm). Austin says that the conservational implicature here is that without the specified condition obtaining I am unable to do the thing in question. In contrast to the indicative, these statements are in the present tense. The unrealised condition on which my being able to swim 5km depends might take some time to realise – 2 months, in the case of the training. Nevertheless, such statements seem to say *something* about *how the agent is now* that grounds the truth of the claim. What makes it true that I could swim 5km if I trained for 2 months is something about how I am now, and what makes it false that Steve could swim 5km if he trained for 2 months is something about how Steve is now.

What goes for 'could' statements also goes for 'could have' statements. When used in the indicative mood, the difference between 'could' and 'could have' statements is that whereas 'could' statements tend to refer to actions in general, 'could have' statements tend to refer to particular occasions on which an action might be performed. Austin describes this by saying that 'could' is *indefinite* whereas 'could have' is *definite* (Austin 1979: 215). The following example illustrates both uses: 'I could swim 5km when I was fifteen, and at 2pm on July 14 1995, I could have swum 5km'. The first part of that statement says that when I was fifteen I used to be able to perform an action of a certain type, namely, swimming 5km. The second part identifies a particular occasion on which I could have successfully swum 5km. Thus, when 'could' is used without the modifier the emphasis appears to be on the agent's having possessed some intrinsic ability or skill to perform actions of that type. When 'could' is used with 'have' as a modifier, it appears that more is being said: the agent didn't just have an intrinsic ability to perform an action of the given type, but also had the opportunity to perform a particular action of that type. Of course, explaining the indicative 'could' in terms of 'able', 'ability' and 'opportunity' is only rough and merely moves the question to what these terms mean. Chapter 2, which involves a study of Vihvelin's recent account of abilities, takes up the issue of abilities in depth. Opportunities are discussed more in chapter 3.

The subjunctive use of 'could have' is a little different. Like the indicative, the addition of 'have' suggests that the speaker has in mind some definite occasion for the performance of a particular action. For example, 'I could have swum 5km at 2pm on July 14 1995 if I had not missed the bus to the pool'. It will be clear from the foregoing that the subjunctive 'could have' is in the past tense, not the present tense as in the subjunctive 'could'. Still, just like the subjunctive 'could', it is plausible to think that the subjunctive 'could have' requires completion by an 'if' clause. This is supported by reflecting on the previous example: the statement 'I could have swum 5km at 2pm on July 14 1995 if I had not missed the bus to the pool' means something like 'I would have been in a position to swim 5km at 2pm on July 14 1995 if I had not missed the bus to the pool'. This is not meant to be an analysis, but something of this form seems to be on the right lines. And the point is as follows: if we remove the 'if' clause from the paraphrase then we're left with the statement 'I would have been in a position to swim 5km at 2pm on July 14 1995'. And as Austin says, this statement seems to be incomplete (Austin 1979: 215). *I would have been in a position to swim 5km*…if what? No such incompleteness is evident with the indicative 'could have': 'I could have swum 5km at 2pm on July 14 1995' becomes, in our rough paraphrase, 'I was in a position to swim 5km at 2pm on July 14 1995' and this does not cry out for any kind of completion.

On the basis of the above distinctions, Austin identified two different claims that one might make concerning the connection between 'could' statements and 'if' statements. First, one might claim that 'could' statements require completion by some 'if' clause in order to have a complete meaning and express a complete thought. As we've seen, this is a claim that is very plausible when it comes to 'could' statements in the subjunctive mood. This might not always be immediately obvious, but that is only because such 'if' clauses are often left unstated, being supplied by context. We could call this claim – that 'could' statements requirement completion by an 'if' clause – the ***completion project***. In contrast to the completion project is what Austin thinks of as the ***analysis project***. This is the claim that the proper analysis of 'could' statements will be in terms of 'if' statements. This claim could presumably be made for 'could' statements in either the indicative mood or the subjunctive mood.

These two projects are very easy to confuse not just because statements containing a 'could' or 'could have' phrase might be either indicative or subjunctive, but also because, as already mentioned, the 'if' clause required by the subjunctive forms (if Austin is right) is often suppressed, being supplied by context. This means that for some sentences involving 'could' or 'could have' it will be impossible to tell, apart from appeals to the wider context, whether the statement is in the indicative mood or the subjunctive. Consider the following example:

> I could have gone to the party

In some contexts this is in the indicative mood. For example, suppose we're discussing Pete's party and you are ruing making a prior engagement that precluded your going. I rub it in your face by pointing out that *I could have gone to the party*. Nothing would have had to have been different in order for me to be able to go: there is no paper I would've had to have completed earlier, no prior engagement I would have had to have cancelled. Just as things were, I could have gone: 'I could have gone to the party' means something like 'I was in a position to go to the party'. In other contexts such an utterance would be in the subjunctive mood: we're both stuck on a late-running train and are ruminating on the things we could've done had the train not been horrendously delayed. Here, my uttering 'I could have gone to the party' is in the subjunctive and means 'I could have gone to the party if the train had not been late'. This cannot be paraphrased with 'I *was* in a position to go to the party' but needs 'I *would have been* in a position to go to the party'.

This confusion is especially easy to make when discussing free will, Austin says (Austin 1979: 215), because such discussions are often conducted exclusively in retrospective terms: we consider only examples in the past tense and so only ever ask whether the agent *could have done otherwise*. This results in those present tense 'can' statements for which we also want an analysis or an account (see below) being excluded from our view, making it easier to conflate the indicative and subjunctive 'could' statements. This in turn leads some to confuse the completion project with the analysis project. Austin charged Moore with making precisely this mistake: he did not clearly distinguish between the completion project and the analysis project and as a result he ran together options (a) (the idea that there are at least three senses of 'could') and (b) (the idea that 'could' statements should be analysed in terms of conditionals) (Austin 1979: 214ff).

Where does all this leave us? My purpose here is not an exegesis of Moore, but to assess the most plausible conditional analysis of the 'can' and 'could' statements relevant to free will, and I will limit my comments accordingly. The first thing to say is that our target is some set of statements in the indicative mood. I take this to be straightforward and uncontroversial but the following can be said in support of it nonetheless.

As recounted in the Introduction, the conception of free will with which we are working is the idea that free will involves the agent having the ability or power to do otherwise. This core idea can be expressed in a number of idioms: in terms of what the agent *can* bring about, what the agent *is able* to do, what the agent *has the power* to do, and so on. It is a substantial question whether anything hangs on the different ways of stating the position and for now I'm taking the various formulations to be ways of getting at the same idea. But each of the three formulations just given are in the indicative mood. And this emphasises the idea that whether an agent has free will depends (in large part) on how things are now with the agent. What matters is whether the agent is – in some sense yet to be articulated – able to do otherwise. What doesn't matter is whether the agent *would* be able to do otherwise *were* some other conditions to be satisfied.

Another way of seeing this is to simply note that our account needs to cover all statements in the present tense pertaining to what an agent can do, including, of course, the word 'can' itself. But 'can' has no proper subjunctive use. It is true that sometimes things like the following are said: 'I can fix that shelf for you; that is, I can fix it if you hire me a power drill'. But such usages are vulgarisms and should more properly be written 'I could fix that shelf for you, if you were to hire me a power drill' (i.e. it is a subjunctive use which is properly expressed with 'could').

Given that our interest is in a certain set of indicative 'could' statements, option (b) – which stated that an 'if' clause is required to complete the meaning of the 'could' statement in question – is ruled out as a way of reading **CA-C**. This leaves us with three initial ways of stating the conditional analysis, **CA-C** as read according to option (a), **CA-S** and **CA-W**. In the following section I will discuss each of these in turn.

1.1.3. Three varieties of the simple conditional analysis

The first of our options, **CA-C**, analyses 'could' with a conditional featuring 'could':

(**CA-C**) 'I could have A-ed' is true if and only if I could have A-ed, if I had chosen to A

This version of the conditional analysis faces two problems which combine to render it implausible. First, the phrase 'could have' in the analysans cannot be the same sense of the phrase as that employed in the analysandum. If it were, then we would immediately encounter a vicious regress: the 'could have' in the analysans would be analysed in its turn by **CA-C**, and so on. So the analysans must be employing a different sense of the phrase. One proposal here would be that the 'could have' in the analysans is the subjunctive 'could have' that Austin was at pains to distinguish from the indicative. That is, the suggestion would be that our analysis of the indicative 'could have' be in terms of the subjunctive 'could have'. These two forms of 'could have' – unlike the indicative and subjunctive forms of 'could' without the modifier – match in tense so there is no problem there. But there is a problem (the second of the problems mentioned above) for such a proposal.

It concerns the nature of the 'if' statement in the analysans. Austin has argued that this 'if' statement is not a genuine conditional but is instead akin to the statement 'there are biscuits on the sideboard, if you want them' (Austin 1979: 215). Austin contends that whatever the nature of these 'if' statements, now often referred to as 'biscuit conditionals' after his example, they are not genuine conditionals: they do not affirm any kind of consequence relation between the two components of the statement. Why would this be a problem? Well, as mentioned above, the motivation of those who propose conditional analyses is to provide an account of the ability to do otherwise that is compatible with the thesis of causal determinism. We saw that the thesis of determinism states that the laws of nature together with the past entail a unique future. This rules out the agent being able to do otherwise *in some sense*. But if the relevant 'can' and 'could' statements are equivalent to (genuine) conditionals then this isn't a problem and the compatibility is achieved. The relevant sense of 'could' would be one which is (roughly) equivalent to a statement such as this: 'if the past had been different such that the agent chose X (instead of Y), then the agent would have A-ed'. Such a statement is straightforwardly compatible with determinism. As Roderick Chisholm summarises the point, 'even if all of the man's actions were causally determined, the man could still be such that, if he had chosen otherwise, then he would have done otherwise' (Chisholm 1966: 15). This reconciliation relies on analysing the 'can' and 'could' statements relevant to free will with (genuine) conditionals. If, therefore, the 'if' statement provided as the analysans is not a genuine conditional the proposed reconciliation will be in serious trouble.

In order to support this position Austin provides two criteria which genuine conditionals satisfy but which other 'if' statements do not: (1) genuine conditionals entail their contrapositive, (2) genuine conditionals do not entail their detached consequent. To illustrate: the conditional 'If I run, I pant' entails the contrapositive 'If I do not pant, I do not run' (or more colloquially, 'If I'm not panting, I'm not running'). It also fails to entail the detached consequent: we cannot infer from 'If I run, I pant' that 'I pant' (or, 'I am panting'). The statement 'If I run, I pant' thus satisfies both tests and counts as a genuine conditional. But 'I could if I chose' fails both of these tests. From 'I could A, if I chose to A' we cannot infer 'If I cannot A, I do not choose to A' (Austin questions whether this contrapositive has any sensible meaning at all). Our 'could' statement also fails the detached consequent test because from 'I could A, if I chose to A' we *can* infer 'I could A' (Austin 1979: 210). In other words, if I could A if I chose, then I could A *whether or not I chose* – my being able to doesn't depend on my choosing.

In attempting to explain what's going on here, Austin notes that our word 'if' descends from words such as 'doubt', 'hesitation', 'condition' and 'stipulation' (Austin 1979: 210). Logicians have focused on the sense in which it means 'condition'. What we find with 'I could if I chose', however, is that one of the other meanings is more prominent. Austin lists a number of related statements which help him to make this point (Austin 1979: 210):

> I could, all I have to do is choose
> I could, but do I choose to?
> I could, but is it reasonable to choose to?

The central idea in each case is that I could do something but there is some measure of doubt or hesitation over whether I will choose, although the precise interpretation depends in addition on the emphasis put on the words. Along similar lines, Kenny has noted that for many verbs, appending 'if I choose' to the end emphasises the fact that the action can be performed *at will* or *on demand* (Kenny 1975: 141). This is most obviously the case with those verbs which express behaviours that for many people are not actions; for example, when the actress says 'I can cry if I choose to' she's emphasising that she can cry on demand, which is just to say that she has the ability to cry. Similarly, Peter Morriss notes that sometimes appending 'if I choose' emphasises that the only thing stopping an agent performing some action was the absence of a choice to do it; for example, 'I could have had you fired, if I'd chosen to' emphasises that I really was able to have you fired, and the only reason I didn't, was that I didn't choose to (Morriss 1987: 64). The similarity between all these 'if' statements, which are in the indicative mood, and the 'if' statement found in the analysans of **CA-C**, suggests that it too (i.e. 'I could have A-ed, if I'd chosen to A') is also in the indicative mood. But if so, then it's not the subjunctive 'could have' that was suggested as a way around the first problem: we are back with an indicative 'could have' in the analysans for which we need to provide an analysis, or else be set upon the vicious regress.

One way of challenging Austin's conclusions here would be to question the validity of his tests for genuine conditionality. D. F. Pears has subjected these tests to close scrutiny and argued that for more complex conditionals the two tests diverge in their results (Pears 1971: 254–5). If this is right then at least one of the tests must be inadequate. This does not falsify Austin's conclusion, however, because even if just one of his tests is correct, the 'could have' statement in the analysans of **CA-C** will not be a genuine conditional. Pears concludes that despite various technical problems with his reasoning, Austin's conclusions concerning these statements were correct. We can safely agree with Robert Kane (1996: 53), then, when he describes the approach embodied in **CA-C** as 'wrongheaded'.

The second suggestion, **CA-S**, is subject to similar problems. Austin states that the occurrence of 'should' here is not functioning as a modal auxiliary but is instead being used to express the speaker's intention (Austin 1979: 210). To illustrate this he presents the following examples which he says are parallel:

> I shall marry him if I choose
> I promise to marry him if he will have me

Austin contends that the whole 'marry him if I choose' clause is governed by the verb *shall* just like the whole 'marry him if he will have me' clause is governed by the verb *to promise*. In other words, Austin takes the former to be an assertion of an intention with conditional content: the speaker intends to <marry him if she chooses>. There is nothing wrong with the idea that intentions might take conditional content. Some have suggested that this is how it is with most of our intentions, even when we affirm an intention to A 'at all costs' (see, e.g., Luca Ferrero (2009: 700)). The problem with Austin's claim is that when the 'condition' is the agent's choosing it appears to negate any attribution of any intention to the agent: 'I shall go to the game if I choose', far from expressing an intention, indicates that I haven't yet formed any intention about the matter. What it seems to assert is something very similar to 'I could go to the game if I chose', namely, that I *can* go to the game, that it's up to me and depends only on my decision. If this is right then **CA-S** is to be handled in the same way as **CA-C**.

By far the most common way of treating the 'should' conditional in **CA-S**, however, is to assume that it was nothing more than a slightly old fashioned way of saying the same thing as the 'would' conditional of **CA-W**, and that Moore meant the same by both (Ferenc Huoranszki's (2011: 55) discussion is an example of this). This is very plausible given Moore's oscillation between the two, and it leaves us with the following analysis which I will refer to as the **simple conditional analysis** or **SCA** for short:

> **Simple conditional analysis**
> (Past) 'S could have A-ed' is true if and only if S would have A-ed, if S had chosen to A
> (Present) 'S can A' is true if and only if S would A, if S were to choose to A

The following sections outline the primary difficulties for this account.

1.2. Problems for the simple conditional account

Discussions of the simple conditional analysis typically classify its failings into two broad kinds: those which argue that truth of the conditional is not necessary for the truth of the 'can' statement and those which argue that the truth of the conditional is not sufficient for the truth of the 'can' statement (see, e.g., Kane 1996: 52–8). This is a useful classification, but what is not often noticed is that there are important distinctions between the different arguments falling into each category. In sections 1.2.1 and 1.2.2 I will consider two arguments which claim that the truth of the conditional is not necessary for the truth of the 'can' statement. The first, by questioning the choice of verb used in the antecedent of the conditional, questions the scope of the analysis. The second challenges the adequacy of 'would' conditionals on modal grounds and so gets at issues concerning the efficacy involved in acting. In section 1.2.3 I will look at a collection of related arguments which purport to show that the truth of the conditional is not sufficient for the truth of the 'can' statement.

1.2.1. The antecedent used in the conditional: choosing, willing, trying

In section 1.1.3 I determined that of the three variants of the conditional analysis offered by Moore, the one involving a 'would' conditional is the most plausible. The past tense version of this simple conditional analysis is as follows:

(**SCA**) 'S could have A-ed' is true if and only if S would have A-ed, if S had chosen to A

We now turn to consider the antecedent of the conditional used in the analysans. Moore employed the verb *to choose* in the antecedent. We saw that when this verb is combined, as part of an 'if' statement, with a 'could' clause in the consequent, there is good reason to think that the result is not a genuine conditional. When the consequent involves a 'would', however, things are different. According to **SCA**, the statement 'I could have walked to the store' is equivalent to 'I would have walked to the store, if I had chosen to walk to the store'. I will assume for the time being that it is straightforward to interpret this as a causal conditional and thus facilitate the kind of reconciliation that Moore is after. With this assumption, *choosing to walk to the store* is to be conceived of as some prior mental event or process in which I resolve to walk to the store and which then causes my walking to the store.

Even with this assumption, however, employing the verb *to choose* in the antecedent is not problem free. There are two kinds of case which threaten the account, both of which begin to illustrate the important role that action theory will play in the current investigation. Consider the following example which is representative of the first kind of case:

> Max runs regularly, his usual route being 5 km. One Saturday his friend Suzie invites him to join her on her weekly running route, a 15 km run. Max has run 15 km twice before, and although it's a lot more than he usually runs, he decides to take up Suzie's offer. Suzie goes easy on him, setting a leisurely pace. Nevertheless, the run is a struggle for Max: he had to exert a lot of effort throughout, and on a few occasions had to give himself a mental pep-talk in order to avoid giving up.

The point of this example is as follows: many of the things we do require more than just our choosing to do them. We don't simply choose, and let nature do the rest; rather, we engage in sustained activity throughout. In the above example, Max does not simply decide to join Suzie on her run and then find himself, an hour and a half later, having run 15 km. Max was active – in a number of different ways – throughout the period of run, and his completing the run depended on all of that activity. The fact that Max might have decided to run 15 km, begun to run 15km, and yet failed, helps to illustrate the point. To put it another way: Max's being able to run 15 km isn't just a matter of some *event* or other occurring as a result of his decision to run 15 km, it is a matter of *a particular kind of event* – namely, a stretch of activity – occurring as a result of his decision. And that stretch of activity includes lots more that Max does and upon which his resulting achievement depends. (Here I've assumed for simplicity that actions and stretches of activity are events; I take it that the thought can be rephrased if, for example, actions are not events, as is argued by Helen Steward (2012)).

Now, this is not how it is for all actions. Sometimes it at least appears to us that we choose something and then find ourselves having done something – as might be the case with some habitual actions, or things that it is possible to do 'on autopilot'. But that there are cases which are like the example above is enough to falsify **SCA** if the antecedent is framed in terms of choosing. Another example, adapted from van Inwagen (1983: 115), helps to make the point more vivid:

> Suppose that Napoleon could have won the Battle of Waterloo. This is not to suppose merely that it was possible *that* he win, which might be made true by, e.g., it having been possible that Wellington drop dead in May 1815. We're supposing here that it was possible *for* Napoleon to win *by something he himself did*. We are supposing, in other words, that there was a course of action that Napoleon had available to him that would have resulted in his victory had he pursued it.

Now, it is a substantial question whether, and how, Napoleon's knowledge of which actions he needs to perform affect what is he able to do. I put this issue to one side for now. In chapter 5 I will discuss different senses of 'able' that can be articulated by taking into account different doxastic and/or epistemic states of the agent. For present purposes, what the Napoleon example illustrates is that if he had won the Battle, he would not have won it just by his choosing to win it, and neither would he have won it by choosing and also exerting more effort in an activity he's already engaged in (as in the running example). If Napoleon had won the Battle he would have done so by engaging in a stretch of activity which included many sub-activities as parts – planning and predicting troop movement on maps, inspiring the generals, rallying the troops, riding here and there on horseback, and so on.

These kind of examples have led some to replace *choosing* with *trying*. The simple conditional analysis would then be written as follows:

> **SCA-Trying**
> 'S could have A-ed' is true if and only if S would have A-ed, if S had tried to A
> 'S can A' is true if and only if S would A, if S were to try to A

Austin (1979) provides some critique of the attempt to use the verb *to try* in a conditional analysis. Chisholm, in a commentary on Austin's discussion, identifies five distinct objections to the view (although Chisholm appears to attribute them to Austin, it is not clear that all are to be found in Austin, and, if they are, Austin did not clearly distinguish them) (Chisholm 1964: 23–4). The five objections are as follows (I have placed them in an order I take to be more natural than Chisholm's):

(1) **There are some things an agent can do, but which cannot be tried**. Chisholm lists closing one's eyes as an example of something which cannot be tried (Chisholm 1964: 23).

(2) **An agent might try to A, but stop due to some interruption**. Chisholm's example is of a man trying to solve a puzzle who is then interrupted by a call to dinner, whereupon he abandons the puzzle. He *can* solve the puzzle, but it was false to say that if he tried to solve it, he would have succeeded.

(3) **An agent might try, but not hard enough**. That is, someone might try to A but only half-heartedly and they might fail as a result. They *can* do it, but only if they try as hard as they can.

(4) **'I can' is consistent with failure due to external interruption whereas 'I would, if I tried' is not**. Austin offered his (now famous) example of the skilled golfer who misses a short putt in support of this. This kind of case differs from (2) in that the failure is due to external interference, rather than altered behaviour on the part of the agent.

(5) **Sometimes an agent is such that there is something he can do, but not if he tries to do it**. This objection is distinct from (1). It's not that the agent *can't* try to perform the action in question (because of the nature of *trying*); rather, the worry is that trying to A will bring about failure to A, but trying to B would in fact bring about success in A. For example, a golfer can hit the ball to point *p* (a precise location), but if he tried to hit the ball to that precise point, he'd (almost certainly) fail.

All of these points aim to show that it is sometimes true to say that an agent can do something but at the same time false to say that they would do it, if they were to try. In other words, the claim is that the truth of the analysans is not necessary for the truth of the analysandum. The strength of these criticisms varies greatly.

Objection (1) is the most straightforward to deal with. Chisholm does not explicitly comment on why it is that an agent cannot try to close his eyes, but there are two things he might be getting at. The idea might be that an agent cannot try to close his eyes because it is too easy, and trying necessarily involves exerting a certain amount of effort to overcome some difficulty. As closing one's eyes is something that is very easy, no effort is expended, and so no trying is involved. Alternatively, the idea might be that an agent cannot try to do anything which can be done 'directly'; anything, that is, which is not done by doing something else. Closing one's eyes is something an agent can do directly in this sense: in contrast to say, entering my house, which I do *by* finding my keys, unlocking the door, opening the door and walking through the doorway, to close my eyes I simply close my eyes.

It might be thought that the proponent of **SCA-Trying** could reply to this first objection as follows. If trying is understood as necessarily connected to the exertion of effort, then it could be pointed out that there are circumstances under which closing one's eyes is difficult. Perhaps one has received an injury to the eye which makes closing it painful and difficult. Or perhaps one suffers from a kind of partial paralysis such that it just takes a lot effort and concentration to close one's eyes. If trying is understood in terms of 'direct' actions, then it might be pointed out that an agent could try to close their eyes by using some kind of mechanism, perhaps one hooked up to the eyelids, such that their action is now mediated by the mechanism. Or, it might be suggested that the agent could close their eyes *by* trying to look as if they were asleep. This latter action is something that can be tried, and so the analysis now covers this kind of action. Chisholm (1964: 24) takes this later route, noting that it requires a small amendment to the analysis:

(**SCA-Chisholm**) 'S can A' is true if and only if there is some B such that if S tried to do B, S would A.

This amended analysis supposedly introduces a disconnect between what you can do and the method by which you do it and so, according to Chisholm, it can now handle the problem cases. But this approach is wrongheaded. Suppose that Abby is in front of us and we are trying to determine whether she can *open* her eyes (I speak of *opening* rather than *closing* because it makes for easier examples; I presume that if Chisholm thinks one cannot try to close one's eyes, one cannot try to open them either). In wondering whether Abby can open her eyes what we want to know is whether opening her eyes is under her control. This might be an open question because, as suggested above, it might be that Abby is recovering from a recent eye injury or that she suffers from some kind of paralysis. But when seeking an answer to the question of whether Abby can open her eyes, it is not to the point to ask whether Abby can open her eyes *by* attempting to look like she normally does when she's awake, or whether she can open her eyes *by* using a mechanism hooked up to her eyelids. Those are different actions and to ask about them is to change the subject. So even if Chisholm is right that Abby can act in mediated way W, and so *try* to act in mediated way W (because all mediated actions admit of trying), and can as a result be said to be able to open her eyes, still, that does not explain why Abby can open her eyes directly or normally (supposing it to be true that Abby can open her eyes normally). In other words, if Abby can open her eyes in the normal way, this is not because, as Chisholm's analysis makes out, Abby can open her eyes indirectly or in a mediated fashion. The existence of these cases – opening one's eyes by using a mechanism attached to one's eyelids, closing one's eyes by trying to look as if you were asleep – is beside the point. Even if Chisholm's analysis successfully handles these deviant cases, it doesn't correctly handle the normal cases.

There is a further reason why this approach is wrongheaded. To try to A is to engage in intentional behaviour which has A as its goal. Brian O'Shaughnessy's definition of trying captures this thought well:

> Trying consists in doing, intentionally and with just that purpose, whatever one takes to be needed if, the rest of the world suitably cooperating, one is to perform the action (O'Shaughnessy 1973: 369).

The point is that to say that someone is trying to A does not say anything as yet about their means. And as O'Shaughnessy's definition makes clear, trying encompasses all the means that are needed. But if that's the case, then when someone is about the business of opening their eyes using a mechanism attached to the eyelids, that person can nevertheless be accurately described as *trying to open their eyes* (as well as *trying to operate the eye-opening mechanism*). This means that Chisholm's amended analysis doesn't get us anywhere because the so-called deviant cases are already handled by the original analysis.

What this illustrates is that Chisholm's statement of the problem is inadequate. We should not object to **SCA-Trying** on the grounds that there could be an agent who can A but who cannot try to A. Rather, we should object that there might be an agent who can only A directly but who cannot try to directly A. For example, imagine that Abby can open her eyes only in the normal way – she has no eye-opening mechanism available. Opening her eyes normally is the only route she has available to opening them. If Chisholm is right that people cannot try to open their eyes in the unmediated way, then Abby can open her eyes, but she cannot try to do so, and so **SCA-Trying** is unable to provide an account of Abby's being able to open her eyes. But now the following will be clear: Chisholm's amended analysis, **SCA-Chisholm**, is of no help with such cases. Chisholm's analysis gives the right result (but for the wrong reason) *only if* we assume that Abby has some non-normal way available to her of opening her eyes (e.g. an eye-opening mechanism); once we preclude such alternative ways of acting, it becomes clear that Chisholm's analysis isn't an advance over **SCA-Trying**.

A far better response for the proponent of **SCA-Trying** is to deny the claim underlying the objection. That is, the proponent of **SCA-Trying** should simply deny that there are any actions which cannot be tried: where there is action, there is trying. There are two kinds of account of action according to which trying is ubiquitous in this way. On the first, *trying* is identified with a kind of mental state – sometimes called a volition or a willing – that is said to precede and initiate all extended bodily movement and behaviour which is properly labelled 'action'. This kind of view is defended by, among others, Jennifer Hornsby (1980) and O'Shaughnessy (1973). There are various ways of spelling out the connection between tryings (conceived of as volitions/willings) and the resulting behaviour. According to Hornsby, the trying or volition is itself the action if it causes (in the right way) some bodily movement. Actions, then, are not preceded by tryings because the tryings just are the actions. This is not essential to the view: one might think of tryings as volitions but identify the bodily movement caused by the volition as the action. Either way, however, wherever there is acting, there is trying. On the second construal, trying is not identified with a distinct kind of event, such as a volition, but is instead identified with the entire stretch of the agent's activity, whatever form it takes. This kind of view is defended by Hugh McCann (1975). My purpose here is not to defend either of these views on the nature of trying. The point is only that adopting one of these views is the best response the proponent of **SCA-Trying** has to the objection above. It does mean that the proponent of **SCA-Trying** is committed to some such view, but the diversity of views according to which trying is ubiquitous seem to preclude this being too much of a problem.

Objections (2), (3) and (4) are similar in the following respect: they each purport to show that the truth of the proposed analysans is not necessary for the truth of the analsyandum because, once the agent has started to act, something might interfere and stop the agent achieving their goal. They will be dealt with in section 1.2.2.

Objection (5) is useful because it highlights the need to distinguish between a sense of 'can' which takes into consideration the agent's knowledge and a sense which doesn't. In what follows I will argue like so: this example only serves as a counter-example if we assume that the golfer does not know which of his actions would lead to his landing the ball at *p*. Once we make that assumption, however, it becomes clear that we are analysing a sense of 'can' which does not require knowledge of what the agent is bringing about. This is problematic inasmuch as it's not yet clear whether (and in what manner) such a sense of 'can' is relevant to free will.

First, to expand on the example a little: Chisholm tells us that the golfer landed the ball at a particular place, *p*. Landing the ball at *p* was, therefore, something the golfer could do. Chisholm appeals to the principle of 'does implies can' to support this claim (Chisholm 1964: 24). However, if the golfer had tried to land the ball precisely at *p*, then 'in all probability' he would have failed. So the agent could have landed the ball at *p* (because he did, and *does* implies *can*), and yet the agent would (almost certainly) have not succeeded had he tried to land the ball at *p*. Thus, **SCA-Trying** fails.

Now to my argument: why, in order to make sense of the example, do we have to assume that the golfer does not know which of his actions leads to landing the ball at *p*? Well, suppose for *reductio* that the golfer does know which action leads to landing the ball at *p*. Perhaps, for example, he knows that he always shoots a little wide. Thus, he knows that in order to land the ball at *p*, he needs *to take aim* at a point two metres to the left of *p*. In such a case, the golfer would be able to land the ball at *p* whenever asked. The golfer would not do this by taking aim at point p; but there is something he could do which would have the desired result. The primary point, however, is that in doing this – taking aim at a point 2 metres left of *p* – the golfer would be *trying to land the ball at p*. So it would be false to say that if the golfer tried to land the ball at *p* he would fail. If the golfer knows which of his actions gets the ball to point *p*, then the statement 'If he tried to land the ball at *p*, he'd fail' is false unless we take it to mean something like 'If the golfer *took aim at point p*, he'd fail to land the ball at point *p*'. This is not an entirely unnatural way of reading 'tried to land the ball at *p*', but that is only because in the usual case, golfers will try to land the ball at *p* by taking aim at point *p*.

What we have to assume then, is that the golfer doesn't know which of his actions would lead to landing the ball at *p*. In this case, it is natural to assume that if he were asked to land the ball at *p*, he would attempt to do so by taking aim at point *p*. This would, according to the details of the story, result in his failure. But Chisholm insists that the golfer can, in this situation, land the ball at *p*. That means that Chisholm must be talking about a sense of 'can' which may be true even when the agent does not know how to bring about the result. He goes on to suggest that cases like this are also handled – like the cases mentioned in objection (1) – with his revised analysis:

(**SCA-Chisholm**) 'S can A' is true if and only if there is some B such that if S tried to do B, S would A.

The idea is that this allows us to affirm, with Austin, that the golfer can indeed place the ball at *p*, and that's because there is something he can do which would have this result, even though he doesn't know which thing it is he needs to do. Chisholm's analysis may well address the problem; that is, Chisholm articulates a sense of 'can' according to which it is true that the golfer can land the ball at *p* (despite not knowing how). What we need to be very clear about is that Chisholm is now presenting an analysis of a sense of 'can' which allows us to say that agents can bring about things which they do not know how to do.

This is not to say that this sense of 'can' is not relevant to free will. It is a substantial question what kind of doxastic or epistemic conditions there are on the sense of 'can' relevant to free will. The important point is that Chisholm nowhere notes this shift in usage, nor does he argue that this sense of 'can' is the sense most relevant to free will. We can draw some further distinctions: in the example as presented it is natural to assume that although the golfer doesn't know quite how to go about landing the ball at *p*, nevertheless, he's got a good idea that there is something he could do to land the ball at *p*. But we could introduce a variant case where the golfer is simply unaware that landing the ball at *p* is a result he can bring about. Suppose landing the ball at *p* requires hitting the ball 400 yards. It might be that the golfer has the strength to hit the ball that far, but it requires swinging the club in a particular fashion. He can swing the club in such a way but he has no idea that this is so. In this variant case, there is a series of steps he could take to hit the ball that far, but he is totally unaware of this. Does Chisholm mean to be providing an analysis of this (even broader) sense of 'can'? More importantly: which one of these is the one we should be interested in when our focus is free will?

Few writers have attempted to articulate a sense of 'can' which incorporates the kind of epistemic criterion seen in the above example. Exceptions include Alvin Goldman (1970: 209ff) who distinguishes between non-epistemic and epistemic abilities, and Morriss (1987: 52–9) who makes a distinction between non-epistemic, epistemic and effective epistemic abilities. I will take up these issues in chapter 5, where I will argue that there are doxastic conditions on the sense of 'can' or 'able' that is most relevant to free will.

1.2.2. Austin: 'I can' is consistent with failure, 'I would, if I tried' is not

The previous section showed that the task of arriving at a suitable conditional with which to equate the 'can' and 'could have' statements is not a trivial one. This section addresses one of the points Austin raised against the conditional analyses of 'can' when framed in terms of trying, namely, the claim that 'I can' is consistent with failure whereas 'I would if I tried' is not. This point is demonstrated by objections (2), (3) and (4).

Objection (2) was that an agent might try to perform some action but then stop due to some unexpected interruption. In such cases, much depends on what happens when the interruption occurs. In Chisholm's example, a man is solving a puzzle when he is interrupted by a call to dinner, at which point he abandons the puzzle. One way of understanding this example is that when the man hears the call, he deliberates about what to do and then makes a decision to abandon the puzzle and proceed to dinner. Understood in this way, objection (2) is similar to objection (3): here the man does not choose to abandon his activity, but simply does not try hard enough in the first place. Neither of these cases poses much threat to the spirit of the conditional analysis, even if they falsify it as formulated in **SCA-Trying**. An agent will (almost) always have the option of abandoning what they are doing before having reached some end; and there is little reason to think that this counts against the agent's being able to perform the action in the sense relevant to free will. Given that such failures are due to the agent's own exercise of control, it seems that it is open to the proponent of the conditional analysis to simply qualify the conditional so as to exclude such cases. The same is true of objection (3): of course an agent might try half-heartedly, but the proponent of the conditional analysis intends to employ a notion of trying whereby the attempt is genuine and a qualification may be made to this effect.

These cases stand in contrast to objection (4), where there is a failure that is not due to the agent's own abandoning of the action. Austin offered his now famous golfing example in support of this point. I quote his original presentation of this objection at length:

> Consider the case where I miss a very short putt and kick myself because I could have holed it. It is not that I should have holed it if I had tried: I did try, and missed. It is not that I should have holed it if conditions had been different: that might of course be so, but I am talking about conditions as they precisely were, and asserting that I could have holed it. There is the rub. Nor does 'I can hole it this time' mean that I shall hole it this time if I try or if anything else: for I may try and miss, and yet not be convinced that I could not have done it; indeed, further experiments may confirm my belief that I could have done it that time although I did not.
>
> But if I tried my hardest, say, and missed, surely there must have been something that caused me to fail, that made me unable to succeed? So that I could not have holed it. Well, a modern belief in science, in there being an explanation of everything, may make us assent to this argument. But such a belief is not in line with the traditional beliefs enshrined in the word can: according to them, a human ability or power or capacity is inherently liable not to produce success, on occasion, and that for no reason (or are bad luck and bad form sometimes reasons ?) (Austin 1979: 218).

Austin's claim is that he could have made the putt *in those very conditions* and assuming that the very best attempt was made – i.e. there was nothing else he *could have done* to improve the chances of his sinking the putt. That is, Austin thinks there is a sense of 'can' which tolerates these kind of exceptions and this must be admitted, he thinks, even if our adherence to other doctrines forces us to accept that there must have been some explanation for the failure.

There has been much commentary on Austin's putting example. I will consider what I think are the three primary responses. The first charges Austin with equivocating on the sense of 'can' he's talking about. When he talks about his being able to make the putt he missed he is referring to a sense of ability or being able to do something closely associated with the notion of having the skill to do it. Austin may well be right that such a notion of ability tolerates exceptions and so will not be analysable in terms of a 'would' conditional. But Austin goes wrong – so the charge goes – in thinking that this is the sense of 'can' being analysed by the proponents of the conditional analysis; rather, they are providing an account of the 'can' relevant to free will and that is not the notion of having enough skill to do something.

This objection is not persuasive. There is no doubt that there is an important distinction between having the power or being able to do something on the one hand and having the skill to do something on the other. But Austin was well aware of this and intended his argument to apply to what he called the 'all-in' sense of 'can' which he characterised in terms of the agent possessing the ability, the opportunity and the skill needed (Austin 1979: 229). Of course, this characterisation of 'can' in terms of ability, opportunity and skill leaves much unsaid, for there are different senses in which someone might be able to do something. Indeed, in chapter 2 I will suggest that the terms 'able' and 'power' can vary in meaning in ways very similar to 'can'. Still, Austin's comments make it clear that, at the very least, he does not take himself to be making the mistake he is accused of here. Austin is claiming, then, that the sense of 'can' or 'able' relevant to free will need not mean that the performance of the action is guaranteed: he was able to sink the putt in the sense relevant to free will, despite his failing when he tried. On the other hand, the truth of the conditional 'if Austin were to try to sink the putt, he would' does indeed require that he succeed given his trying. Thus the relevant 'can' statement cannot be analysed by the conditional.

Bernard Berofsky (2012: 75) agrees with this point – i.e. that Austin is not guilty of shifting focus to a sense of skill – but has argued that Austin's worry is irrelevant. Indeed, Berofsky suggests that all objections to the conditional analysis which are based on the showing that the truth of the conditional is not necessary for the truth of the 'can' statement are irrelevant. This is because, Berofsky suggests, the compatibilist does not need to provide an *analysis* of the relevant sense of 'can' or 'able'. Many compatibilists – with Moore as the prime example – have indeed attempted to provide an analysis of the relevant sense of 'can'. But to do so would be to do more than what is needed in order to vindicate compatibilism. According to Berofsky, all the compatibilist needs to do is show that the truth of the conditional is sufficient for the truth of the relevant 'can' statement, not that it is also necessary. He puts the point like this:

> Although we present [the simple conditional analysis] as a biconditional, compatibilism succeeds on a merely conditional interpretation. The compatibilist need only insist that, if—but not only if—an agent were to succeed if she were to try to A, then she could have done A (Berofsky 2012: 75).

This is not a new point. David Sanford highlighted it in a 1991 paper where he said the following:

> Moore's argument for compatibilism does not require an analysis of *could have*. What matters is implication in just one direction. With this example, the argument for compatibilism goes through so long as 'I would have sunk the putt if I tried' implies 'I could have sunk the putt' (Sanford 1991: 209).

Sanford doesn't consider the point very important because he thinks that other arguments – those to be discussed in the next section – show that the one-way implication also fails. In the next section I will agree with Sanford on this point. But there is also reason to doubt that compatibilism requires *nothing more than* a one-way implication between the 'would' conditional and the 'can' statement. Suppose that Berofsky's claim is correct: the truth of 'if S were to try to A, then S would A' suffices for the truth of 'S can A'. Concede too that this sense of 'can' is relevant to free will. This doesn't yet secure compatibilism unless it is also assumed that there is no other sense of 'can' relevant to free will which doesn't also admit of some compatibilist analysis. More precisely, Berofsky's objection requires one of two assumptions: either he must assume that there is just one sense of 'can' or 'able' that is relevant to free will, or he must assume that there are other senses of 'can' relevant to free will but that they too admit of compatibilist analyses. Establishing this would not be easy, as is demonstrated by the variety of distinctions that have so far been identified for further discussion. I take it that Berofsky's point blunts Austin's objection to some degree but is not as conclusive as he considers it.

1.2.3. Could the agent have tried to A?

A third objection to the simple conditional analysis articulated, or at least hinted at, by a number of authors (C. D. Broad (1952); Chisholm (1966); Lehrer (1968)), challenges the idea that truth of the conditional is sufficient for the truth of the 'can' statement. Consider the following example from Lehrer:

> (**Red Candy**) Suppose that I am offered a bowl of candy and in the bowl are small round red sugar balls. I do not choose to take one of the red sugar balls because I have a pathological aversion to such candy. ... It is logically consistent to suppose that if I had chosen to take the red sugar ball, I would have taken one, but, not so choosing, I am utterly unable to touch one (Lehrer 1968: 32).

It is not essential that the example is framed in terms of choosing; it would be just as problematic if it were put in terms of trying. Lehrer's point is that his pathology could rule out the choosing, the trying, or whatever the proponent of the simple conditional analysis puts in the antecedent. In such a situation Lehrer cannot take the candy, but the conditional 'if Lehrer tried [choose] to take the candy, he would' is true and so the analysis must be false. Note that this doesn't just refute the conditional account when presented as an *analysis* of 'could' statements, i.e., when presented as a biconditional. It refutes the account even if the compatibilist is relying only on a one way implication from conditional to 'could' statement. In other words, this is a problem for the simple conditional account even if Berofsky's point discussed at the end of the previous section is correct.

Chisholm has presented a similar example. Consider a man, S, who shoots another. It might be the case that if S had chosen to do otherwise then he would have done otherwise but also the case that, for whatever reason, he couldn't so choose – perhaps he has a pathology as in Lehrer's example (Chisholm 1966: 16). In such a case the conditional is true, but this appears to make no difference to what S could do: if he couldn't *choose* otherwise, he couldn't *do* otherwise.

In discussing their respective examples, Lehrer and Chisholm share a similar approach: they both try to find a set of statements which are consistent with the conditional proposed by the advocate of the simple conditional analysis but which together entail the negation of the 'can' statement. And their examples share a similar structure. Nevertheless, Lehrer and Chisholm explain the problem differently. Indeed, Lehrer's explanation of the problem does not fit the structure of his red candy example. It fits instead the structure of a second example that Lehrer gives which runs as follows:

> (**Paralysis**) Suppose that, unknown to myself, a small object has been implanted in my brain, and that when the button is pushed by a demonic being who implanted this object, I became temporarily paralyzed and unable to act. My not choosing to perform an act might cause the button to be pushed and thereby render me unable to act (Lehrer 1968: 32).

Lehrer's idea here is that at some particular time t, he doesn't choose to perform some act. This causes the demon to paralyse him, thus removing Lehrer's ability to perform the act. But it could still be true that if Lehrer choose to perform the act, he would. This is because the demon might decide to remove Lehrer's paralysis as soon as he makes the choice to act. Thus, at t it's false that Lehrer can perform the act, but it's true that were he to choose to perform the act, he would. Were he to so choose, he would be given back the ability to perform it (by the demon). This kind of example concerns what has been called, in the literature on dispositions and powers, a *fink*: a factor which is triggered when someone without an ability to A attempts to A and which will bestow the ability onto the person quickly enough for the ability to A to be successfully exercised. Cases involving finks have been discussed a great deal in the dispositions literature and I will deal with these cases in the following chapter, where I discuss Vihvelin's dispositionalist account of free will – an account which explicitly treats the ability to do otherwise as a matter of possessing certain dispositions (see 2.1.3).

For the remainder of this chapter I will put such cases to one side and focus on the kind of cases exhibited by Lehrer's **Red Candy** case and Chisholm's shooting example. I will call the problem, as understood by Chisholm, the **Transfer Problem**, because at the core of the complaint is the idea that an impossibility to choose or to try transfers over to an impossibility to perform an action. I will label the 'would' conditional that the proponent of the conditional analysis intends to use as the analysans **WC** and will call the 'could' statement being analysed **M** (because it's the modal claim relevant to free will). One thing to be aware of: Chisholm put something like the following argument forward in a number of places (Chisholm 1964; 1966), but Bruce Aune (1967: 191) has reported that Chisholm misstated his case, at least in the (1966) article, and tells us that Chisholm communicated to Aune a revised or corrected version of the objection he envisaged. I'm working with this revised version. I will use Lehrer's **Red Candy** example rather than Chisholm's shooting example because it concerns an action that was not performed, and thus avoids having to introduce a number of complications to do with the nature of refraining (refraining is discussed in chapter 5). In **Red Candy** Lehrer is such that he cannot choose to take a red candy due to some pathology, but it is nevertheless true that he would have taken one if he had chosen. This gives us the following:

(**M** – analysandum) Lehrer could have taken a red candy
(**WC** – analysans) If Lehrer had chosen to take a red candy, Lehrer would have taken a red candy

Chisholm claims that the reason why examples of this kind falsify the simple conditional analysis is because the following two statements are consistent with **WC** but together they imply the falsity of **M**:

(**C1**) Lehrer would not have taken the red candy, had he not chosen to take the red candy
(**C2**) Lehrer could not have chosen the red candy

C1 makes the assertion that the only way for Lehrer to take the red candy is by choosing to do so. **C1** together with **C2** implies:

(**~M**) Lehrer could not have taken the red candy

But if **C1** and **C2** are consistent with **WC** and if they imply **~M** then the 'would' conditional, **WC**, is also consistent with **~M** and it therefore cannot be the correct analysis of **M**. Indeed, there cannot even be a one-way entailment from **WC** to **M** as suggested by Berofsky. If Chisholm is right about **C1** and **C2** entailing **~M** then his larger argument is sound; it relies on the fairly uncontroversial idea that if P, Q and R are consistent, and P and Q entail ~S, then R cannot entail or be logically equivalent to S. The crucial thing is whether the entailment from **C1** and **C2** to **~M** goes through.

Aune has attacked Chisholm's argument at just this point; this inference, Aune claims, is suspicious because it involves modal claims. Aune starts by re-construing **C1** as equivalent to:

(**C3**) If Lehrer had taken the candy, Lehrer would have chosen to take the candy

He then concedes that there is *some* sense of 'could' on which the inference from **C1** and **C2** (or, now, **C3** and **C2**) to **M** goes through:

> …[there is], after all, the familiar modal principle that if P entails Q and it is impossible that Q, then it is impossible that P. Assuming that 'could not' may mean 'it was impossible', this principle may thus warrant the inference that if S's choosing [to fire the shot] is in some way necessary for S to [fire the shot] and if S cannot, in some sense, choose to [fire the shot], then, in that same sense of 'cannot', S cannot [fire the shot] either (Aune 1967: 192).

The difficulty for Chisholm, Aune thinks, is that he needs to assume that this inference goes through for the sense of 'could' relevant to free will. This is 'highly questionable' because since 'the time of Moore, philosophers have generally agreed that not every sense in which a man cannot do something is equally relevant to the question of his freedom' (Aune 1967: 192). Aune establishes this latter point with the following kind of example: suppose a woman's choosing to work for another hour is causally sufficient for her working for another hour and that the woman does indeed choose to work for a further hour. Then, given that she has chosen to continue working, any other action will be 'relatively impossible' (Aune 1967: 192).

Aune takes this kind of impossibility to be harmless and thinks that the sense in which S cannot fire the shot is of this same harmless variety. To think otherwise is to rely on the questionable assumption that the sense in which one could have *done* otherwise is the same sense in which one could have *chosen* otherwise (Aune 1967: 193). Aune thinks this is clearly mistaken because willing and choosing are not voluntary actions and so couldn't possibly be *performed* in any sense at all (Aune 1967: 193).

There is much that is puzzling in this critique of Chisholm. Aune starts by establishing two things which purportedly help him to rebut Chisholm's attack. First, he points out that there is more than one way in which an action might be impossible. I take it that this claim is straightforward. Second, he says that some of the ways in which an agent *cannot* do A are harmless and do not affect an agent's freedom. This claim is supported by his causally sufficient willing example: if the woman wills to work for another hour, and her willing to work is causally sufficient for her working, then any other action is 'relatively impossible'. In other words, any other action is impossible given causally sufficient antecedents; any other action is causally impossible. These first two points are, I take it, correct, but it's unclear how they are meant to help in blunting Chisholm's point. Suppose we agree, in addition to these two points, that the sense of 'cannot' employed by Chisholm is contested. Still, *all* Aune has shown is that *some* senses of 'cannot' are harmless. But if that's all he has shown, then at the very most he has established a dialectical stalemate. In other words, on the assumption that Aune's two initial claims are relatively straightforward, we are still left with the substantial question whether the sense in which Lehrer's taking the red candy is impossible is the sense relevant to free will. Aune does nothing to settle this matter; nor does he provide any considerations which might lead us to think that advocates of the **Transfer Problem** are focusing on a sense of 'can' that is not relevant to free will. Aune has merely emphasised that there are different kinds of impossibility; but that was already well known.

It is to be doubted however whether Aune has even succeeded in establishing a stalemate because his example of a harmless impossibility is not analogous to Chisholm's case. In Aune's case the woman cannot perform a different action (read: overt, bodily action) given the choice that *she herself has made*. That is, the 'cannot' in Aune's example is harmless only because we assume (a) that the woman's choice to continue working was itself was under her control, and (b) that her continuing to work flows from the choice she made. Moreover, it is *only* true that the woman cannot perform a *different overt bodily action*; Aune's example is not one where the choice itself is something that the woman could not have refrained from.

In **Red Candy** type examples, however, the impossibility affects more than the overt bodily action which flows from the choice. It affects the choice itself. Lehrer is unable to make the choice and therefore unable to act: the impossibility transfers from the choice to the action. There is therefore a world of difference between the two cases: in Aune's example the woman controls her choice whereas in **Red Candy** Lehrer is not in control of what he chooses. Aune, therefore, has so far failed to make his case.

Aune says one final thing in support of his response to Chisholm's objection. Chisholm's case, he says, rests on the assumption that 'the relevant sense in which a man cannot *do* otherwise may be the same as the sense in which he cannot *choose* to do otherwise' but that this is to be doubted because 'for most philosophers, choosing and willing are not voluntary actions' (Aune 1967: 193).

Two things need to be said in response. First, Aune's conclusion does not follow from his premiss. Suppose Chisholm accepts that the sense in which someone chooses otherwise is different to the sense in which someone does otherwise (we can even suppose he agrees to this because he thinks, like Aune, that willing and choosing are not voluntary actions). Still, it might be that there is a sense of 'can' and 'cannot' which, although not the sense of agency, entail something about the 'can' relevant to agency. For example, it seems plausible to suggest that if it is metaphysically impossible for an agent to perform some action, then the agent will be unable to perform that action, in the sense relevant to free will, whatever that sense of 'able' turns out to be. Metaphysical impossibility, in other words, constrains what agents can or are able to do. And we can agree with that even if we don't yet know what the relevant sense of 'able' is. Something similar might be said of causal impossibility: perhaps it too constrains what an agent is able to do in the relevant sense, and perhaps we could affirm this even without knowing what the relevant sense of 'able' is. This would be a more contentious claim, but it might be argued for by noting that humans are physical beings and that agency is a causal phenomenon. I do not wish to defend this line of thinking here; I raise it only to point out that Aune's conclusion does not follow straightforwardly from his premiss. If he wants to establish that the impossibility associated with Lehrer's being unable to choose to take the red candy is unproblematic because choices are not voluntary actions, he needs to say a lot more about why this is so.

Second, Aune's premiss is questionable. The claim that most philosophers think willing and choosing are not voluntary actions was contentious even when Aune was writing, a time when there was a general air of hostility towards even the existence of *willings* and *choices* (Hardie 1971). Van Inwagen felt able to express his puzzlement, in 1983, over why anyone would question the idea that choices are actions (van Inwagen 1983: 236 n.11). Today it is a commonplace to think that choices are actions and a number of powerful arguments can be marshalled in support of this claim (See, e.g., Thomas Pink (1996: 3-5, 66-75); Richard Holton (2009: Ch 3)). If choices are actions then they are things we *do* or *perform* and Aune's emphasis of the difference between things we *choose* and things we *do* dissolves.

A very different kind of solution to the **Transfer Problem** is to accept the basic point but attempt to repair the analysis by including in it an assertion to the effect that the agent can try or choose to perform the action. In other words, the suggestion is that to say that Lehrer could have taken the red candy is to say that if Lehrer had chosen to take the red candy, he would have done so *and that Lehrer can choose to take the red candy*. This suggestion is not straightforward because if we include in the analysans a statement affirming that the agent can choose or try to perform the action in question, then we are, prima facie, including an instance of the analysandum for which we are attempting to provide an account.

There are two ways the proponent of the conditional analysis might attempt to handle this worry. The first is to suggest that this second occurrence of 'can' means something different to the 'can' which we are analysing. Van Inwagen (1983: 116–9) has shown that pursuing this line of thought is problematic. Let us refer to the second sense of 'can' as 'CAN'. Then van Inwagen's key insight is that there is a condition on the meaning of 'CAN' that must be fulfilled if the approach is to succeed, namely, that statements of the form 'S CAN A' must entail the corresponding statements of the form 'S can A'. If this entailment doesn't hold then it will still be possible to produce counter-examples similar to Lehrer's **Red Candy** case; indeed, **Red Candy** itself will still serve as a counter-example. The adding of the 'Lehrer CAN try to take the candy' assertion will do nothing to address the concern if that assertion doesn't entail 'Lehrer can try to take the candy'. Not only does finding a sense of 'CAN' which has this feature seem to be just as difficult as the original problem, it might be suspected that it just is the original problem. After all, if an account of 'CAN' satisfying this constraint could be given, then that account could just be applied directly to the original 'can', without any intermediary step.

The second option available to the proponent of the conditional analysis is to attempt to find some antecedent for which the question *But* can *the agent try [choose] to A?* does not arise. The obvious choice here is to cite the agent's *wanting* or *desiring* to A. This would give us something like the following account:

> (**SCA-Desire**) 'S can A' is true if and only if S would A if S were to want to A

Van Inwagen thinks that this goes some way to solving the problem because wanting A, unlike choosing A or trying to A, is not an action. We could still ask whether the agent could have wanted to do something different. But if we do ask this question we won't be asking whether there was something the agent could have *done*; rather, we will be asking whether it was possible for the agent to have had a different want or desire. Van Inwagen thinks this is significant. If the agent is unable to choose to A then it seems as if A-ing is not under their control. On the other hand, the same is not true if the agent is unable to want or desire to A. Van Inwagen says this is because 'people sometimes do what they have no particular desire to do' (van Inwagen 1983: 118).

The analysis rendered in terms of wants and desires faces its own problem, however: sometimes possessing a different desire alters what an agent is able to do. Consider again Lehrer's red candy case. Lehrer has a pathological aversion to red candy. Presumably this means that when offered some candy he will experience no desire to take one. Indeed, he will very much want to avoid taking one. But the conditional 'If Lehrer wanted a red candy, he would take one' might still be true because if Lehrer did possess such a desire he wouldn't have his pathology. Van Inwagen says there are more straightforward cases: suppose that Smith is in a coma. Given that he's in a coma he cannot get out of bed. But it could still be true that if Smith wanted to get out of bed, he would. If Smith had that desire, he'd no longer be in a coma. Van Inwagen thinks this issue can be solved with the following analysis:

> (**SCA-Van-Inwagen**) x could have done y =def if x had wanted to do y, x would have done y and x's wanting to do y would not have been sufficient (in the broadly logical sense) for x's possessing an advantage with respect to doing y that x did not actually possess (van Inwagen 1983: 120).

In addition to the conditional, this analysis stipulates that the possession of the desire must not be *broadly logically sufficient* for x's possessing any *advantage* which would aid x in doing y. The notion of an advantage is introduced to solve the comatose type cases: sometimes, in order to assess what would happen when someone possesses a desire, we have to change various things which facilitate the possession of that desire. Some of these changes, however, are illegitimate. If someone is in a coma, then in order to desire to perform some action we need to suppose that person is no longer in a coma: the person's desiring to perform an action is broadly logically sufficient for that person's not being in a coma. But as the person is in a coma, this is illegitimate. The complexity here arises because sometimes a person's possessing a desire to perform some action results in the possession of an advantage with respect to the performance of that action, but this advantage is something the agent acquires via acting. For example, my wanting to get to New York might lead me to ring up the airline and book a seat on a flight to New York. Having a seat on a flight to New York is an advantage when it comes to getting to New York. So we can't simply stipulate, when formulating the analysis, that the possession of the desire is the only thing different. We need to allow advantages that the agent acquires via acting, but exclude those advantages that must be posited if the agent is to have the desire. Van Inwagen's requirement that the desire not be *broadly logically sufficient* for any advantages not already possessed is designed to achieve this (van Inwagen 1983: 120).

The notion of an advantage was introduced into the analysis of 'can' by Lehrer, whose account will be considered in chapter 4. Van Inwagen's discussion is useful because his suggestion that the advantages not be broadly logically sufficient promises to solve a problem which plagued Lehrer's account, as we will see. Van Inwagen thinks this succeeds as a conditional account; he doesn't think it establishes compatibilism because he thinks the conditional employed in the analysis of 'can' just is the compatibilist's central premiss (van Inwagen 1983: 121). I will argue that Van Inwagen's account is unsatisfactory, however, and needlessly concedes too much. This is because it does not solve the problem underlying the Lehrer/Chisholm counter-examples. (Lehrer's (1976) own account, which first introduced the notion of an advantage, is not a conditional account).

The problem with van Inwagen's account is his contention that moving to an antecedent which is not an action solves the problem raised by the Lehrer/Chisholm style counter-examples. He is surely right that if we switch to a conditional where the antecedent involves the agent's wanting to A rather than choosing or trying to A, and if we then ask whether the agent *could have wanted to A*, we're not asking about whether the agent *could have performed an action*. We must be asking, as he says, whether it is possible – in some as yet unspecified sense – that the agent had had a different want (van Inwagen 1983: 118). What I want to suggest, however, is that this is not the crucial point. The **Transfer Problem** is a problem because being unable to choose or try to A means that the agent is unable to A. Choosing to A is the only way that the agent can A; or at least, choosing to A is the only way the agent can A freely. If the agent can no longer choose to A, then A-ing is no longer under that agent's control. Van Inwagen says that this isn't the case with desiring or wanting: the agent may not be able to want to A but may still be able to A because, as previously quoted, 'people sometimes do what they have no particular desire to do'. In other words, even if a person has no desire to A – and even if there is no possibility of that person desiring to A – they might still be able to A.

But this point depends on adopting a particular conception of desire. The sense in which it is uncontroversial to say that people sometimes do things they have no particular desire to do has been called, by G. F. Schueler, the 'ordinary sense' of desire (Schueler 1995: 1–2). This is the sense in which we can think about and reflect on our desires, the sense in which we may not identify with all of our desires, and may have, for example, a second-order desire to change some of our first-order desires: I don't want to eat the salad for lunch, but I might want to want to eat the salad for lunch. In this sense, as van Inwagen emphasised, I can – I'm able to – eat the salad even though I do not want to. But **SCA-Desire** (and van Inwagen's strengthened version, **SCA-Van-Inwagen**) will not be plausible if 'want' is understood as employing this sense of desire. This is most easily seen with examples where someone has the desire but doesn't act: I desire to eat fudge every day, but I don't because it's unhealthy. Given that my desire to eat fudge does not result (very often) in my eating fudge, according to **SCA-Desire** I am unable to eat fudge.

If **SCA-Desire** (and van Inwagen's strengthened version, **SCA-Van-Inwagen**) are to be plausible they will have to be understood as employing what Schueler calls the 'philosophical sense' of *want* or *desire*. This is understanding might be glossed as 'wanting, all things considered'; on this conception, *wanting* or *desiring* are now closely associated with *trying to get* and, on many views of what it is to act, this notion of desire will be connected analytically to action (Schueler 1995: 1–3). That is, it will always be correct to say of someone who A-s that they wanted – in this 'philosophical sense'; in the 'all things considered' sense – to A. The problem is that on this sense of desire it will no longer be true – or at least, it will be much less plausible – to say that people sometimes do what they have no particular desire to do. The assumption that Chisholm made about choosing – namely, that if S doesn't choose to fire the shot, then S won't fire the shot – now seems to hold true of desire: if S doesn't want (in this philosophical 'all things considered' sense) to fire the shot, then S won't fire the shot. On this conception, without a desire or want to A there will be no A-ing. Rendering the simple conditional analysis in terms of wants or desires, therefore, does not escape counter-examples of the form of those provided by Chisholm and Lehrer. I conclude, then, that the **Transfer Problem** is decisive against the simple conditional analysis, even in its more sophisticated forms. In the following chapter I will look a recent attempt to present a contemporary version of the conditional analysis by appealing to recent work in the metaphysics of dispositions.

Chapter 2 – Vihvelin's dispositional compatibilism[2]

2.1. Vihvelin's account

2.1.1. Introduction

The previous chapter assessed the prospects for the simple conditional analysis of the 'can' of free will. The simple conditional analysis is characterised by two things: (a) the use of a single 'would' conditional in the analysans, and (b) making no mention of the source of the behaviour which occurs (it does not, for example, assert that the behaviour is produced by the agent). In this chapter I will consider Vihvelin's account of free will which is an attempt to repair the simple conditional analysis.

Vihvelin finds in Moore two claims. First is the claim that agents have abilities in virtue of having dispositions. Second is the claim that those dispositions admit of a simple conditional analysis. Vihvelin says correctly that these claims are independent and that we can accept the first whilst rejecting the second. Vihvelin suggests that the objections usually thought to refute Moore's account – those surveyed in the previous chapter – only threaten the second claim. Moreover, she does not think that these objections are anywhere near as decisive against this second claim as they are usually thought to be. So although Vihvelin concedes that the objection concerning finks (see 1.2.3, and below 2.1.3) has some force, she thinks that the objection I deemed decisive against the simple condition analysis (which I called the **Transfer Problem**) is flawed. This assessment is part of what allows her to advocate a repair of the conditional analysis rather than its abandonment.

Section 2.1.2 outlines in broad strokes Vihvelin's view on free will. Here I introduce her distinction between skills, narrow abilities and wide abilities, a distinction that Vihvelin considers crucial to a proper understanding of free will. In section 2.1.3 I present in more detail a key part of Vihvelin's account, namely, her account of intrinsic abilities (which she calls narrow abilities). I then move to some assessment. In section 2.2 I argue that Vihvelin's analysis is, just like the simple conditional analysis, decisively refuted by the **Transfer Problem**. In section 2.3.1 I raise a problem for Vihvelin's account of intrinsic abilities (her narrow abilities). This problem undermines her account of the ability to do otherwise. I demonstrate this (in the following chapter) with reference to her treatment of the Frankfurt-style cases.

2.1.2. A broad overview of Vihvelin's three notions of ability

In introducing her understanding of free will, Vihvelin says the following:

> A central part of our view of ourselves as agents with free will is the belief that we are able to choose from a diverse array of possible courses of action. [We] care about our possibilities, both with respect to our immediate choices and with respect to the overall shape of our lives. ... [And] we take for granted

[2] Portions of chapters 2 and 3 were previously published in Kittle (2015). Thanks go to Springer Netherlands for allowing inclusion here.

> that we have a genuine choice about what we do, that what we do is not the only thing we are able to do, that it is in our power to choose and act in ways other than the ways we actually choose and act, that how we choose and act is in our control and up to us (Vihvelin 2013: 1).

Vihvelin does not follow van Inwagen in *defining* free will as the ability to do otherwise, but she does say that 'the belief that we have the ability to do otherwise is conceptually central to our commonsense view of ourselves as free and responsible agents' (Vihvelin 2013: 6). One of Vihvelin's aims is to argue that we do often have this freedom that we ordinarily think we have, and, moreover, that this freedom is compatible with determinism. One of the main obstacles to both of these projects is understanding what sense, or senses, of 'able' are employed by this commonsense understanding of free will. Vihvelin therefore begins by articulating three different senses of 'able'.

It is worth noting that with this framing of problem Vihvelin has already departed from the terms of Moore's discussion. That discussion and the debate which followed was framed in terms of 'can'. But as Vihvelin notes, although 'can' may be used to say that an agent is able to do something – and this is how Moore was using the word – it may also be used to express various notions of possibility that have nothing to do with agency. This move is paralleled by van Inwagen in his recent writings. He says the following:

> Many philosophers, in attempting to spell out the concept of free will, use the phrase 'could have done otherwise'. I did so myself in *An Essay on Free Will*. Nowadays, however, I very deliberately avoid this phrase. I avoid it because 'could have done otherwise' is ambiguous and (experience has shown) its ambiguity has caused much confusion in discussions of free will. … [It] sometimes means 'might have done' … and sometimes 'was able to do' (van Inwagen 2008: 332).

Undoubtedly, the focus should be on issues which pertain to the agent's control. But the move is not entirely innocent. First, few writers have attended to differences between 'ability' and 'able'. Both van Inwagen and Vihvelin, for example, proceed to treat 'ability' as nothing other than the noun form of 'able'. Thus we see Vihvelin take it for granted that if an agent has an ability to A, then there is a sense in which that agent is able to A; and if there is a sense in which the agent is able to A, then the agent has a corresponding ability to A (Vihvelin 2013: 7–8). But this is not straightforward for the terms appear to have different, even if overlapping, meanings. Morriss reminds us of this by drawing our attention to the relevant definitions from the *Oxford English Dictionary* (Morriss 1987: 81; my emphasis):

> **Ability**:
> 2. The quality *in an agent* which makes an action possible; suitable or sufficient power (generally); faculty, capacity (to do or of doing something).
>
> **Able**:
> 4. Having the qualifications for, and means of, doing anything; having sufficient power (of whatever kind is needed); *in such a position* that the thing is possible for one; qualified, competent, capable.

The pertinent phrase with respect to 'ability' is 'in an agent' whereas with respect to 'able' it is 'in such a position that the thing is possible'. In other words, whilst 'ability' is most naturally used to refer to one of the agent's intrinsic properties, 'able' naturally expresses the idea that something is possible for an agent in virtue of their current circumstances. This can be seen with the Olivia example presented in the Introduction: Olivia is sat in her office when I sneak up and lock the door from the outside. There is a sense of 'able' according to which it is natural to say that Olivia is able to leave her office and also a sense where it is natural to say that Olivia is unable to leave her office; with 'ability' this is not the case: it is natural to say Olivia has the ability to leave her office, but it is much more strained to say that Olivia does not have the ability to leave her office (because of the locked door).

Morriss even discusses senses of 'able' which attribute only the existence of what we might intuitively think of as an opportunity, to the exclusion of any intrinsic abilities. So we might say, for example, that so-and-so is, in virtue of being a British citizen, able to apply for a British passport. And this might be affirmed even if the person does not have the abilities required to apply for a passport; if, for example, the person cannot read or write any English. I will not address these complications, but will instead follow Vihvelin and van Inwagen in treating 'ability' as nothing more than the noun form of 'able'. This simplifies things greatly; but in taking this route it must be remembered that no assumptions can be made about the intrinsic or extrinsic nature of the properties thereby ascribed.

Vihvelin recognises three senses of 'able'. The first expresses the idea of a *skill* or *competence*. When you learn Chinese or how to play the piano, you gain a skill (or more likely, a set of skills), and we can describe this by saying that you now have the ability to understand Chinese or the ability to play the piano. Vihvelin characterises skills as highly stable properties of persons that persist through fairly substantial changes in both the person and their environment. For example, you don't lose the skill to understand Chinese when you're not listening to Chinese; nor do you lose it when you don't have access to any Chinese books; nor when you go to sleep or get drunk. The skill persists through these changes in you and your environment. Vihvelin identifies skills with the notion that some philosophers have called 'general abilities', although she finds the latter notion too vague to be useful (Vihvelin 2013: 7).

Following on from the idea of a skill is Vihvelin's notion of a *narrow ability*, which is glossed as being a matter of 'having what it takes' to perform some action. Possessing the narrow ability to A entails possessing the skills required to A, but it also requires that the agent has 'psychological and physical capacity to use those skills' (Vihvelin 2013: 11). An example illustrates what Vihvelin is getting at here: suppose Bridget and Colin are equally good cyclists. It is midday and Bridget is awake, alert and fully in control of her faculties. Colin, by contrast, has just finished off an entire bottle a vodka and as a result can hardly stand up. Both Bridget and

Colin have the skill to ride a bike. Each of them *can* or *is able to* ride a bike in that each possesses the skill. Nevertheless, there is clearly a sense in which Bridget is able to ride a bike but Colin is not: this sense would be emphasised if we were to add a time index, e.g., Bridget is able to ride a bike *right now*. As Vihvelin understands it, the difference between them is that Bridget possesses the narrow ability in addition to the skill, whereas Colin only possesses the skill. Vihvelin says that narrow abilities are intrinsic properties and as a result they conform to the following supervenience thesis:

> (**Intrinsicness Thesis**) Necessarily, if two persons are intrinsic duplicates governed by the same laws, they have exactly the same narrow abilities (Vihvelin 2013: 13).

Vihvelin's third sense of 'able' introduces and incorporates the idea of an *opportunity*. Sometimes in ordinary language we *contrast* an ability with an opportunity. In such cases the terms 'able' and 'ability' may refer either a skill or to a narrow ability. For example, if I say 'you could've made that shot, if only you had the ability', I'm emphasising (a) that you had the *opportunity* to make the shot (circumstances were right) and (b) that you did not have *the skill* to make the shot. And if Bridget were to reproach Colin, saying 'you could've come on the bike ride, if only you'd been able', she would also be emphasising (a) that Colin had the *opportunity* to ride a bike but (b) that Colin did not have the *narrow ability* to ride a bike (as opposed to denying that Colin had the skill to ride a bike).

Other times, however, we use 'able' and 'ability' to include the presence of an opportunity. Return again to the example of Olivia sitting in her office. Before I sneak up and lock the door Olivia has the narrow ability to leave her office as well as the opportunity to do so. Her surroundings are, as Vihvelin puts it, 'favourable'. After I lock the door Olivia loses the opportunity to leave: her surroundings become unfavourable. Vihvelin calls the sense of 'ability' which reflects the presence or lack of an opportunity *wide ability*. So in the Olivia example, when I lock the door Olivia loses her wide ability to leave her office, whilst retaining her narrow ability to leave (Vihvelin 2013: 169–70). Wide abilities are therefore extrinsic properties.

Vihvelin says that an agent has an opportunity to A when the surroundings are 'friendly', and she glosses 'friendly surroundings' as those which do not contain any extrinsic factors that would stop the agent from exercising their narrow abilities. Olivia's surroundings, for example, are not friendly with respect to her narrow ability to leave her office (Vihvelin 2013: 174, 209).

Vihvelin, after delineating these three senses of 'ability' and 'able', puts forward the following thesis:

> (**ABD**) To have an ability to act is to have a disposition or bundle of dispositions (Vihvelin 2013: 171).

This thesis is intended to apply to both skills and narrow abilities. As a thesis applied to skills it is relatively uncontroversial. It is common to treat skills as dispositions or capacities, and it is widely accepted that skills are compatible with determinism and that the 'can' or 'able' associated with skills is not the 'can' or 'able' relevant to free will (Vihvelin 2013: 174). Vihvelin's substantial claim is that **ABD** applies to narrow abilities. It is a substantial claim because it is not at all clear that abilities are equivalent to dispositions. Moreover, when combined with an account of dispositions framed in terms of conditionals, it will facilitate what Vihvelin calls an 'ontological reduction of free will' (Vihvelin 2013: 170).

Van Inwagen is one of those authors who resists this assimilation. He writes the following:

> The concept of a causal power or capacity would seem to be the concept of an invariable disposition to react to certain determinate changes in the environment in certain determinate ways, whereas the concept of an agent's power to act would seem not to be the concept of a power that is dispositional or reactive, but rather the concept of a power to originate changes in the environment (van Inwagen 1983: 11).

To illustrate the idea of a causal capacity, van Inwagen gives the example of someone who can *understand* French: if the person were to hear French being spoken, they would (definitely) understand what was being said. Someone who can *speak* French, on the other hand, has an ability: there need not be any set of circumstances under which such a person would (definitely) speak French. The ability or power to speak French seems to be under the agent's control in a way that the disposition or capacity to understand French is not.

Vihvelin thinks this reasoning is faulty because it incorrectly assumes that the conditions which trigger an exercise of some disposition – the so-called *stimulus conditions* – must be 'determinate states of the object's environment'. This assumption is problematic because it overlooks those dispositions which have stimulus conditions that are internal to the object which possesses the disposition (Vihvelin 2013: 172–3). Vihvelin argues that it is plausible to think that there are such dispositions. For example, cats are disposed to seek food in response to the internal trigger of hunger, humans are disposed to sweat when they get too hot and radioactive particles are disposed to decay without the need for any trigger whatsoever (Vihvelin 2013: 172).

Once we recognise that dispositions can have stimulus conditions that are internal to the object which possesses the disposition, thesis **ABD** becomes much more plausible, Vihvelin thinks. This plausibility increases still further once we start to get clear about just which particular abilities and dispositions constitute our free will. This is where the *bundle* aspect of Vihvelin's view becomes important. To have an ability is to have a disposition *or a bundle of dispositions*. And because abilities are of the same ontological kind as dispositions, this allows us to say that abilities might be made up of further abilities. Vihvelin goes on to say that 'a highly interesting subset of our narrow abilities' are constituted by the intrinsic disposition to do A in response to *one's trying to do A* (Vihvelin 2013: 175). This is the case for actions as diverse as walking in a straight line and deliberating over what to do. I have the narrow ability to walk in a straight line because I am disposed to walk in a straight line in response to my trying to walk in a straight line; I have the narrow ability to speak English because I'm disposed to speak English when I try; and I have the narrow ability to make up my mind because I'm disposed to do so when I try. The abilities just mentioned will be made up from abilities to weigh reasons, form intentions, perceive one's environment, move one's limbs, and so on; in turn, these latter abilities can be understood in terms of simpler abilities or dispositions.

Given the role that *trying to A* is playing here it is worth asking whether an agent needs the narrow ability to try to A in order to possess the narrow ability to A. As Vihvelin (2013: 175–7) herself recognises she seems to face a dilemma here: if she says that agents do not need the ability to try, then the complaint might come that narrow abilities provide no more control over an agent's actions than mere dispositions do, which is to say, very little control. Just as water does not control whether or not it dissolves salt, people with nothing more than the disposition to do A when they try to do A (i.e. those *without* any ability to try to do A) do not control whether or not they A. On the other hand, if Vihvelin says that agents do need the ability to try to A in order to have the ability to A then she's open to the threat of a regress: does the agent not then also need the narrow ability to try to try?

Vihvelin refuses to consider this a genuine dilemma. She says that 'having a disposition to give response R to stimulus S doesn't entail having any additional disposition to give S as a response to some prior stimulus' (Vihvelin 2013: 176).[3] As such there is no regress and the answer is simple: an agent does not need the narrow ability to try to A in order to have the narrow ability to A. It follows from this that an agent does not need the wide ability to try to A in order to have the wide ability to A. In saying this, Vihvelin is careful not to say that the ability to try is an incoherent idea – rather her point is simply that an agent might possess the ability to A without also possessing the ability to try to A. I will return to the issue of trying in 2.1.4 and 2.2.

2.1.3. Vihvelin's account of intrinsic abilities

This section considers in more depth Vihvelin's account of intrinsic abilities, i.e., those she calls narrow abilities. Vihvelin's account treats abilities as dispositions, so I will begin with a review of the recent literature on dispositions. This will provide the context for an in depth discussion of Vihvelin's account, and will also be useful in the next chapter.

Dispositions are properties which concern how objects behave. They are modal properties, similar in kind to abilities, capacities, powers, tendencies and so on. Indeed, much of the philosophical literature makes no distinction between these kinds of property and uses the term 'disposition' to refer to all of them. I will follow this practice for now; in the following chapter we will see how various differences can be marked out.

[3] See also Vihvelin (2013: 200).

Like the philosophical discussion of ability, discussions of dispositions in the twentieth century began by attempting to analyse such concepts in terms of conditionals. And just as there was a simple conditional analysis of the 'can' of free will (see section 1.1.3), so there was a simple conditional analysis of 'is disposed to'.

(**SCA-Disposition**) 'O is disposed to M when S' is true if and only if, if O were S, then it would M

In this is schema, 'O' is to be replaced by the object; 'M' stands in for the type of behaviour which will result if the disposition is fired, typically referred to as the *manifestation*; and 'S' is to be replaced by the set of *stimulus* or *triggering conditions* – those conditions which initiate the manifestation of the disposition. Most philosophers today consider this analysis to be fatally undermined by two phenomena: *finks* and *reverse finks*. The idea of a fink is simple enough: a disposition is finked when the stimulus conditions, which would normally cause the disposition to manifest, in fact causes the disposition to be lost so quickly that it cannot manifest. Reverse finks are the opposite: the object doesn't have the disposition, but when the stimulus conditions obtain the object gains the disposition, which then manifests. The original examples come from C. B. Martin and according to David Lewis, despite only being published in 1994, they date from the early 1970s (Lewis 1997: 143). The following is adapted from Martin's original example:[4]

> Suppose a live wire – a wire *disposed to conduct electricity when connected to a conductor* – is connected to a machine, an *electro-fink*, which can provide itself with reliable information as to exactly when the wire it is connected to is touched by a conductor. When such contact occurs the electro-fink reacts by making the wire dead for the duration of the contact. It does this so quickly that no electricity is conducted. In other words, the electro-fink ensures that the wire is live when and only when it is *untouched* by a conductor.
>
> Example: At time t1 the wire is untouched by a conductor. By hypothesis the wire is live at t1. But the conditional 'if the wire were touched by a conductor then electrical current would flow from the wire to the conductor' is false of the wire at t1 because the wire's being touched would result in its disposition being lost.

The wire's being live is a disposition: the disposition to conduct electricity when connected to a conductor. The simple conditional analysis states that dispositions can be analysed using a single, counterfactual conditional. In this example the associated conditional would be 'if the wire were connected to a conductor, it would conduct electricity'. However, the presence of the *electro-fink* means that whenever this disposition is put to the test the wire will become dead and so not conduct any electricity. By hypothesis the wire is live and so is disposed to conduct electricity but the counterfactual which is purported to be an analysis of that disposition is false. Alexander Bird summarises the phenomenon like this: 'a finkish disposition is one which is made to go

[4] Martin's original example is found in (1994: 2–3). However, he introduced the notion of a fink by beginning with a *finkish lack of disposition*, which tends to confuse matters. Also, the conditional he cites in his original paper is an indicative conditional, not a subjunctive. I have adapted the example to account for these points.

away by the same stimulus as the stimulus to which the disposition is a disposition to respond' (Bird 1998: 227).

Martin also described the reverse situation: a dead wire might be connected to an *electro-fink* machine now operating on a 'reverse cycle'. The electro-fink now ensures that if the wire is ever connected to a conductor it will become live. By hypothesis the wire is dead but due to the presence of the electro-fink the wire will gain the disposition to conduct electricity if it is connected to a conductor. In this case the wire has a *reverse fink*.

Finks and reverse finks mean that the truth of the conditional is neither necessary nor sufficient for the presence of the disposition. When a disposition is finkish the object possesses the disposition but the associated conditional is false: the truth of the conditional is thus not necessary for the object's possessing the disposition. When an object has a finkish lack of disposition the object does not possess the disposition but the associated conditional is true: the truth of the conditional is therefore not sufficient for the possession of the disposition.

According to Lewis (1997: 147), finks and reverse finks teach us that dispositions are intrinsic properties. When a disposition fires, the manifestation is caused by some of the object's own intrinsic properties. But we can always arrange circumstances such that the firing of the disposition is interrupted due to the disposition being lost before the result occurs; that's what happens with finks. What we need to do, therefore, is stipulate in the analysis that the intrinsic properties of the object which contribute to the production of the manifestation are retained. Lewis thus amended the simple conditional analysis as follows:

> **Reformed Conditional Analysis**
> Something x is disposed at time t to give response r to stimulus s **if and only if**, for some intrinsic property B that x has at t, for some time t' after t, **if** x were to undergo stimulus s at time t and retain property B until t', **then** s and x's having of B would jointly be an x-complete cause of x's giving response r (Lewis 1997: 157).

This account stipulates that the object or agent possesses an intrinsic property B which is retained for some period of time and which forms part of the cause (along with the stimulus conditions) of the manifestation. The notion of an *x-complete cause* derives from John Stuart Mill's idea of a *total cause*. Just as the *total* cause is the *total* set of conditions needed to obtain for some effect to occur, so the *x-complete* cause is the set of all the conditions required, *so far as x is concerned*, to bring about the result (Lewis 1997: 156). With this amendment Martin's cases no longer serve as counter-examples to the analysis because the conditional, now more complex, requires that the property underlying the disposition or ability be retained and this is not the case in the electro-fink examples.

Many writers agree that Lewis's analysis successfully solves the problem of finks. However, there are a number of similar phenomena that also cause problems for conditional analyses, even Lewis's reformed analysis. Mark

Johnston (1992) describes cases of *masking* and *mimicking,* which parallel finks and reverse finks. In cases of masking an object has a disposition but when the stimulus conditions occur the disposition does not manifest *despite its being retained*. Johnston gives the example of a glass which has been safely packaged: the glass is disposed to break when struck, and it retains this disposition, but when it is struck, it doesn't break because it has been encased in foam packaging. And as Michael Fara has recently commented, masks are as commonplace as safely packaged glassware:

> [The] dispositions of objects are being masked all the time. I'm disposed to go to sleep when I'm tired; but this disposition is sometimes masked by too much street noise. Cylinders of rubber are disposed to roll when placed on an inclined plane; but this disposition can be masked by applying a car's brakes. A piece of wood in a vacuum chamber is no less disposed to burn when heated than is its aerated counterpart; but wood won't burn if heated in a vacuum (Fara 2005: 50).

In cases of mimicking, an object displays behaviour which appears to be the result of a disposition but which is in fact caused by something extrinsic to the object. Johnston gives the following example:

> A gold chalice is not fragile but an angel has taken a dislike to it because its garishness borders on sacrilege and so has decided to shatter it when it is dropped. Even though the gold chalice would shatter when dropped, this does not make it fragile (Johnston 1992: 232).

Faced with such counter-examples, many have suggested that we need to be more careful about how we specify the disposition. Lewis takes this view. He advocates a two-step solution to the analysis of any dispositional concept. First, we have to specify the stimulus and the manifestation correctly so as to exclude masks. Second, we apply the **Reformed Conditional Analysis**.

The first step is non-trivial and it is important to see that it applies both to those dispositional concepts for which we have coined terms, so-called *covert dispositional terms* such as 'poisonous', 'fragility' and 'solubility', and also to simple but overt descriptions of dispositions such as 'is disposed to crumble' and 'is disposed to commit crimes'. The latter are often called *overt dispositional locutions*. Lewis thinks that both covert dispositional terms and simple, overt dispositional locutions need to be 'correctly specified' so as to exclude masks *before* the **Reformed Conditional Analysis** can be applied (Lewis 1997: 153). For example, although we might initially think that *being poisonous* should be characterized as 'is disposed to cause death when ingested', this is in fact mistaken. Such a characterisation would only be a 'rough start' and we would need to amend it in order to cope with situations where, for example, someone ingests a poison and simultaneously ingests its antidote (Lewis 1997: 153). Lewis himself does not pursue this thought any further. He simply suggests that once we have the final characterisation we will then be able to apply his **Reformed Conditional Analysis**.

There are a number of different ways that this 'getting specific strategy', as David Manley and Ryan Wasserman have called it, might be carried out (Manley and Wasserman 2008: 63ff). Unfortunately for Lewis, however, none of them are successful. First, we might try to get specific by *explicitly excluding the problematic cases*. Consider a fragile glass. We might initially characterise its fragility as 'is disposed to break when struck', but we would quickly realise that we need to exclude cases where the glass is safely packaged. This would lead us to 'is disposed to break when struck and not safely packaged'. The first point to be made here is that this will not do unless we can explicate 'safely packaged' without appealing to the idea of being disposed to break. That is, if 'safely packaged' here just means 'packaged such that it won't break when struck', then the proposed translation of 'fragile' is 'is disposed to break when struck and not packaged such that it won't break when struck'. This is not informative. So in excluding obstacles we need to be a bit more careful, as with the following: 'is disposed to break when struck and not packaged in 3cm of foam'. Of course, this is only the beginning. We need to exclude the glass being protected not just by foam packaging, but by bubble-wrap, feather cushions, corrugated cardboard, and so on. And we need to exclude the glass being struck with a balloon, a soft toy hammer, and so on. This task looks like it will be unachievable for there are an infinite number of possible interfering factors and we cannot exclude them all. Commenting on this approach, Stephen Mumford says the following:

> The possible interfering background conditions cannot be excluded in a finite list that is appended to the conditional. This is because there is no finite list that could name all such possible conditions in which the manifestation is prevented. [And to] state that the excluded background conditions are any conditions which interfere with the disposition manifestation is to render the conditional trivial (Mumford 1998: 88).

Most agree with this conclusion and so have tried different approaches. Manley and Wasserman (2008) have investigated the possibility of getting specific by focusing on the positive characterisation of the conditions in which the behaviour *is* to be expected. They write that the Lewisian might formulate something like the following as the translation of 'fragility' ('SD' stands for Specific Disposition):

> (**SD**) N is disposed to break when dropped on Earth from one metre up onto a solid surface with a Shore durometer measurement of 90A, through a substance with a density of 1.2 kg/m^3 (Manley and Wasserman 2008: 63).

The idea behind this proposal is that by specifying determinate values for many of the key details of the case we might be able to list conditions which ensure the manifestation occurs. But as Manley and Wasserman note, while this example succeeds in excluding many masks, there are some that would still pose a problem. Indeed, Johnston's original example of a glass which is wrapped in packaging would still pose a problem (Manley and Wasserman 2008: 64). This is problematic, but Manley and Wasserman think there is a deeper worry. Suppose we accept, for the sake of argument, that we have constructed a highly specific characterisation of a disposition along the lines of **SD** and that it does exclude all potential masks. How does such a characterisation, Manley and Wasserman ask, relate to the dispositions we actually ascribe? Manley and Wasserman take Lewis's idea to be that our ordinary ascriptions of dispositions express such specific dispositions:

> Lewis's proposal about the ordinary term 'poison' is that it actually expresses a precise dispositional property, though articulating its stimulus conditions may not be a trivial task. While we may have thought that the stimulus condition for poison is simply being ingested, or that the stimulus condition for fragility is being struck, these are actually only approximations of the real stimulus conditions (Manley and Wasserman 2008: 64).

This same thought, Manley and Wasserman convincingly argue, is also found among those who pursue a third strategy for solving the problem of masks, namely, the appeal to some set of 'ordinary', 'normal', or 'ideal' conditions. So for example, we might, with Sungho Choi (2008), suggest that 'fragility' can be understood as the disposition to break when struck *in ordinary conditions*. Or we might follow Wolfgang Malzkorn (2000) in appealing to *normal conditions* or Mumford (1998: 87–92) in appealing to *ideal conditions*.[5] Whether we appeal to the notion of ordinary, normal, or ideal conditions, it is clear that the set of circumstances posited cannot simply be *that set of circumstances which ensures breaking* (Manley and Wasserman 2008: 64). For *everything* is such that it is disposed to break when dropped under circumstances which ensure breaking. Neither can these circumstances be understood as *conditions which in no way interfere with the manifesting of the disposition*. That would risk triviality in the same way that understanding 'safely packaged' to mean 'packaged such that it won't break when struck' did. So to posit such a set of circumstances must be to posit some (rather precise) set of circumstances which are not analytically connected to the manifestation of the disposition and in which the behaviour would occur, should the stimulus be received. With Choi's ordinary conditions this content comes from our concept of the relevant disposition (Choi 2008: 813–4). Something similar is true of Malzkorn's normal conditions, although he also says that normal conditions may need to be filled out an empirical science (Malzkorn 2000: 457–8). What matters for the present point, however, is not where this content comes from but just that it is posited. For Manley and Wasserman think that any view which attempts to solve the problem of masks by positing such highly specific content – this includes both the second and third approaches – and which also uses conditionals in the characterisation or analysis of the disposition, will face a dilemma.

The dilemma runs as follows: are highly specific dispositions such as **SD** to be associated with a conditional which stipulates that the circumstances must *match exactly* the details of **SD** or a conditional which says that

[5] Mumford's view differs from that of Choi's and Malzkorn's because it is a realist account; still, if one seeks to elucidate the identity of a disposition using a conditional, as Mumford does, then the problem remains. Indeed, I will argue below that there is in any case a parallel problem to be faced by most realist accounts.

the stimulus circumstances *may match or exceed* the values specified? Put another way, when it comes to the conditional used to analyse **SD** we have a choice between the following (Manley and Wasserman 2008: 66–7):

> (**Exact**) N would break if dropped on Earth from ***exactly*** one metre onto a surface with a Shore measurement of ***exactly*** 90A, through a substance with a density of ***exactly*** 1.2 kg/m^3.

> (**Interval**) N would break if dropped on Earth from ***over*** half a metre onto a surface with a Shore measurements ***greater*** than 50A, through a substance with a density ***less than*** 50 kg/m^3.

There are problems whichever route we take. Consider **Exact** first. The proposal that **SD** should be analysed using **Exact** is falsified by a special kind of masking case that Manley and Wasserman have dubbed *Achilles' heel cases* (Manley and Wasserman 2008: 67). In these cases an object does not have a disposition but is such that it would produce the behaviour consonant with the manifestation in some very narrow set of circumstances which are, crucially, paradigm test conditions for the disposition. Manley and Wasserman ask us to consider a sturdy brick. The brick is not fragile and can be thrown around and bashed about without any sign of it cracking or breaking. But it has a weak spot, an *Achilles' heel*, and if it is dropped onto this weak spot from just the right height and just the right angle the brick will break. Significantly, we cannot remove the possibility of these cases by getting even more specific. The highly precise set of circumstances in which the brick would break when dropped might well be a paradigm case for testing fragility. This gives us a recipe for finding counter-examples to any proposed analysis along the lines of **Exact**. For example, with fragility, whatever precise circumstances are appealed to – whatever determinate values each parameter takes – there will be some brick that has a weak spot which matches those conditions. By hypothesis, that brick is not fragile, but according to the analysis being proposed it would be.

Now turn to **Interval**. This proposal seeks to analyse dispositions using a conditional which employs some interval of values. Manley and Wasserman canvas a number of ways of interpreting **Interval** but they think the most plausible is as follows (Manley and Wasserman 2008: 68–9):

> If N were dropped from some height or other over half a metre, N would break.

The idea is 'that something has the disposition to break if dropped if and only if it is such that it would break if dropped in one or another of the circumstances within the specified intervals' (Manley and Wasserman 2008: 70). On the standard possible worlds treatment of counterfactuals, this requires that the object break in the closest possible world(s) in which it is dropped in a paradigm case. But this proposal fails. The sturdy brick with a weak spot causes trouble once more: if it is dropped just right in a world which is a paradigm test case then the closest world(s) will be one(s) where it breaks.

Manley and Wasserman take this problem to count decisively against the conditional analysis. But they go on to raise three further problems which they take to challenge the very idea of giving a conditional analysis

(Manley and Wasserman 2008: 71–4). I will briefly introduce the three problems here; more will be said about them in the remainder of this chapter and the next.

First is the problem of *comparative dispositional ascriptions*. Dispositions, it would seem, are *gradable* in that we can say things like 'glass A is more fragile than glass B'. Any account which analyses fragility in terms of a single conditional will run into problems here because the truth of a single counterfactual is an all or nothing affair. Second is the problem of accounting for the context-dependency of dispositional ascriptions. Manley and Wasserman think they have established 'that "fragile" expresses a different property in the mouth of the aerospace engineer than it does in the mouth of the chemist' because what counts as fragile is context-dependent. They relate this issue to the former: intuitively, whether something counts as fragile or not depends on *how fragile it is* and whether its level of fragility meets some context-specified threshold. Third is the problem of absent stimulus conditions: some dispositions appear to have no specifiable stimulus conditions because they have no stimulus conditions at all. Someone might be highly disposed to talk and yet there be no particular kind of situation that elicits this response: they are just generally talkative (Manley and Wasserman 2008: 72). A similar thing might be true for other character traits: irascible, irritable, miserable, sociable, and so on. Plausibly, the dispositions of radioactive particles to decay might also lack stimulus conditions.

In response to these problems Manley and Wasserman counsel a different approach. Vihvelin incorporates a key part of their approach into her final account of abilities so it is important to see how it works. The primary lesson to take away from the failure of conditional analyses, according to Manley & Wasserman, is that when an object has a disposition to M there is no guarantee that it will M in *all* of some suitably demarcated set of cases. Indeed, it might not even be that an object disposed to M when S will M in *most* of the circumstances where S obtains. To cope with this, they propose the following schema:

> (**PROP**) N is disposed to M when C if and only if N would M in some suitable proportion of C-cases (Manley and Wasserman 2008: 76).

To explain: 'N' is the object, 'M' the manifestation and 'C' the stimulus conditions. Manley and Wasserman are clear that C describes an *event type* of some generality; a C-case, on the other hand, fills out the details of how that event type might occur. For example, if we consider the disposition to break when dropped, then *being dropped* is the stimulus condition, C. But objects can be dropped in different ways and in different environments. In an individual C-case each aspect of the situation that affects whether the manifestation occurs needs to be fully specified with a determinate value. Each C-case therefore describes just one way that an event of type C might occur. As Manley and Wasserman put it, a C-case 'is a fully specific scenario that settles *everything* causally relevant to the manifestation of the disposition' (Manley and Wasserman 2008: 72;

my emphasis). Moreover, only those C-cases which have the same physical laws as the actual world are counted as relevant.

One way of understanding their proposal is as follows: we can imagine each use of their schema to involve quantification over the entire range of C-cases. Their claim would then be that an object possesses the disposition if the behaviour ensues in some 'suitable proportion' of those cases. Rather than associating a disposition with a counterfactual, then, Manley and Wasserman directly quantify over the C-cases. For example, suppose that fragility is accurately represented by the overt dispositional locution 'is disposed to break when dropped'. Their account states that:

> (**PROP-Fragile**) N is disposed to break when dropped if and only if N would break in some suitable proportion of dropping-cases.

On the current interpretation we should take their view to involve quantification over the entire range of C-cases. The range of possible C-cases might begin like this:

> (**C-case-1**) N is dropped on Earth from **exactly 1.0 metres** onto a surface with a Shore measurement of exactly 90A, through a substance with a density of exactly 1.2 kg/m^3.
> (**C-case-2**) N is dropped on Earth from **exactly 1.1 metres** onto a surface with a Shore measurement of exactly 90A, through a substance with a density of exactly 1.2 kg/m^3.
> (**C-case-3**) N is dropped on Earth from **exactly 1.2 metres** onto a surface with a Shore measurement of exactly 90A, through a substance with a density of exactly 1.2 kg/m^3.
> ...

Obviously these descriptions of particular C-cases (based on Manley and Wasserman's toy example) are simplified greatly. Each C-case is supposed to be a fully specified way that the object N could be dropped. So each C-case will be a complete description of all the factors which might affect whether the manifestation occurs. Still, I think the idea will be clear enough. Why do Manley and Wasserman only require that the behaviour occur in a 'suitable proportion' of such cases (as opposed, say, to most such cases)? The reason is that they think the threshold will vary according to the disposition in question and context. For example, water's disposition to dissolve salt intuitively involves salt dissolving in a very high proportion of cases: the disposition is, almost, invariable. As Van Inwagen might put it, water dissolves salt 'willy-nilly'. On the other hand, to count as irascible – disposed to get angry – I do not need to fly into a rage at every opportunity. I might count as irascible if I get unduly angry on only three occasions out of every ten, say. There is also contextual variation: in the mouth of a building site labourer 'fragile' will pick out a different property to that which it picks out when uttered by the owner of a jewellery store. Manley & Wasserman's account has the potential to accommodate both kinds of complexity.

This is a significant move by Manley & Wasserman. They are doing a lot more than merely denying that the connection between the stimulus and the manifestation can be understood in terms of a single counterfactual.

To begin with, they are denying that there is just one value that 'a suitable proportion' may take for each given stimulus and manifestation pair. On their view, different objects are disposed to break to when dropped *to different degrees*. And in different contexts, what counts as a 'suitable proportion' of the dropping cases varies, such that which objects we label as fragile changes with the context. This means that on Manley and Wasserman's view, underlying our ascriptions of fragility is a set of modal properties of the form *disposed to break when dropped to degree X* (Manley and Wasserman 2008: 72). Objects have these modal properties regardless of the context, and regardless of the threshold currently associated with the term 'fragile'. This means that Manley and Wasserman are not proposing anything like the reductionist account that Lewis had in mind. It also means that the degree to which an object is disposed to exhibit some behaviour has now become part of the definition or characterisation of the disposition. For example, the brick and the glass are both disposed to break when dropped, just to different degrees; the only difference between the brick's disposition and the glass's, is the degree to which it is disposed to break. So this 'degree of breakability' is now part of the definition of the disposition. To put it slightly differently, if we are told that an object is disposed to break when dropped, we do not have a full understanding of that property until we're told to what degree it is disposed to break when dropped. Intuitively, this feature of dispositions is the 'strength' of the disposition. Barbara Vetter (2010: 21), borrowing a term from Angelika Kratzer's (1981) semantics, calls this the *modal force* of the disposition.[6] I will employ this terminology in what follows.

With that background now in place we can turn to Vihvelin's account. In her earlier work Vihvelin developed an account of abilities which closely tracked Lewis's reformed conditional analysis of dispositions (See, e.g., Vihvelin (2000; 2004)). In her more recent work she has conceded that the problem of masks – both the original masks and the sophisticated Achilles' heel cases – cause a problem for her early account. However she does not adopt Manley and Wasserman's account of dispositions, for she thinks it has a number of undesirable features. First, it does not insist that dispositions are intrinsic properties – something which Vihvelin thinks is a mistake. Second, it quantifies over all possible stimulus condition cases which share the same laws as the actual world. Vihvelin (2013: 185) thinks this is implausible for epistemological reasons (outlined below).

As a result, what Vihvelin aims to do is to combine her earlier Lewisian account with the insight from Manley and Wasserman that she takes to be key to addressing the problematic counter-examples. This key insight is the idea that the modal force need not be any form of restricted necessity (as it is with all accounts that employ a single counterfactual). This results in the following ('LCA' stands for **L**ewis's **c**onditional **a**nalysis; 'PROP' was the name of Manley and Wasserman's account of dispositions):

[6] Note that while Vetter borrows the term from Kratzer, the concept she applies it to is different; Vetter (2015: 70) argues, for example, that a variable modal force is just what is needed to explain the gradability of dispositions, whereas Kratzer attempts a different kind of explanation.

> **LCA-PROP-Ability**
> S has the narrow ability at time t to do R in response to the stimulus of S's trying to do R **if and only if**, for some intrinsic property B that S has at t, and for some time t' after t, **if** S were in a test-case at t and S tried to do R and S retained property B until time t', **then** in a suitable proportion of these cases, S's trying to do R and S's having of B would be an S-complete cause of S's doing R (Vihvelin 2013: 187).

This account closely follows the Lewisian structure that Vihvelin has previously employed; thus 'B' here refers to the intrinsic property that Lewis posits as the disposition's causal base. The key insight from Manley and Wasserman is seen in that Vihvelin requires only that the agent, S, perform an action of type R in a 'suitable proportion' of cases. There is however a significant departure from Manley and Wasserman's account at this point. Manley and Wasserman say that a disposition is possessed if the manifestation is seen in a suitable proportion of *all* C-cases which have the same laws as the actual world. This means that to consider whether an object is disposed to break when dropped, we have to consider what would happen in every possible scenario (sharing our laws) where that object is dropped. This includes, as Vihvelin says, cases where the object is dropped in the middle of a tornado, under water, on far off distant planets, and so on. Vihvelin thinks this is implausible because we just don't seem to have that kind of modal knowledge (Vihvelin 2013: 185). Thus, Vihvelin's account doesn't quantify over all of the stimulus condition cases (C-cases) which share our laws; instead, it quantifies over just those C-cases which share our laws *and which are in some way significant to the disposition in question*. Such C-cases Vihvelin calls *test-cases*.

How does Vihvelin determine what the makes a C-case a test-case – put otherwise, what does this *significance* amount to? It is here that Vihvelin appeals to the 'getting specific' strategy: recall that this strategy was one approach to solving the problem of masks which involved making the disposition's stimulus conditions and manifestation more specific. Vihvelin agrees that such a strategy cannot solve the problem of masks *on its own*. But what it can do, for any given stimulus and manifestation pair, is determine some fairly narrow set of relevant test-cases. This will exclude some – perhaps the majority of – masking cases. The remaining cases will then be handled by requiring only that the manifestation behaviour be produced in some 'suitable proportion' of cases (rather than in all the test cases). Vihvelin's use of the 'getting specific' strategy will be explored in more depth below.

This concludes the presentation of Vihvelin's account of narrow abilities, via a thorough but crucial discussion of the recent literature on dispositions. In the following subsection (2.1.4) I will answer the following question: according to Vihvelin, which kind, or kinds, of ability are relevant to free will? Then in the subsequent two sections (2.2, 2.3) I will raise a number of problems for Vihvelin's account.

2.1.4. Which kind (or kinds) of ability is (are) relevant to free will?

We are now in a position to ask the following question: which kind (or kinds) of ability does Vihvelin think is (are) needed for free will? Unfortunately, Vihvelin does not answer this question directly. Although she wants to refrain from defining free will as including the ability to do otherwise, she does affirm that 'the belief that we have the ability to do otherwise is conceptually central to our commonsense view of ourselves as free and responsible agents' (Vihvelin 2013: 6). This statement is made before she draws her distinction between narrow and wide abilities, but it's clear from comments made later in the work that she takes this commonsense view of ourselves to affirm the presence of wide abilities. She says, for example, that '[t]here is something at the heart of our commonsense way of thinking about ourselves as agents with free will ... which does appear to assume that we are at least sometimes in situations in which we have the wide ability to do otherwise' (Vihvelin 2013: 14). This is what our everyday experience of choice involves: the belief that we often have the ability to deliberate, decide, and do otherwise. Moreover, Vihvelin is clear that she thinks this everyday picture is often veridical. It's not just that we *believe* we often have the wide ability, it's often the case that *we really do have* the wide ability to deliberate, decide and do otherwise (Vihvelin 2013: 168, 192-3).

What complicates her account is that Vihvelin thinks a person might possess free will without possessing 'the free will we think we have' (Vihvelin 2013: 110–4). In other words, to have free will on a particular occasion, it's not necessary that the agent have the wide ability to deliberate, decide and do otherwise. Does Vihvelin think that free will always requires wide abilities of some sort or other? Or perhaps sometimes narrow abilities alone suffice? We get help at this point from a distinction Vihvelin makes between three different kinds of choice: **Moorean Choice**, **Shackled Choice**, and **Owellian Choice**. The details are as follows:

> (**Moorean Choice**) Ordinary, everyday choice situations. In such situations, an agent who faces a decision between two courses of action, A and B, 'really is able to decide to do either' and would successfully act on the basis of whatever decision is made (Vihvelin 2013: 113).
>
> (**Shackled Choice**) An agent in a shackled choice situation is physically restrained such that 'he cannot move any part of his body' (Vihvelin 2013: 114). But he is not subject to any mental pressure, indoctrination, brainwashing and so on. As such, although the agent is unable to perform a different overt bodily action, the agent can perform more than one mental action. Political prisoners (who are tightly constrained) are the paradigm examples of this kind of situation.
>
> (**Orwellian Choice**) An agent in an Orwellian choice situation is subject to extreme physical pressures (e.g. beatings, starvation, torture) as well as severe mental abuse (e.g. indoctrination, brainwashing, etc). The example Vihvelin cites here is Winston Smith, from Orwell's *1984*, during his time at the Ministry of Love. Consider Winston when he is lying immobile on a gurney and being given electric shocks whenever he gives the 'wrong' answer. According to Vihvelin, in such a situation Winston is still able to *try* to defend his beliefs; 'the range of [his] freedom and choice has been circumscribed to a very small radius, yet it is clear that he still has all the freedom and all the choice that is necessary for moral responsibility. He remains able to defy O'Brien, at least momentarily. He remains able to try or begin to say, or at least to think, the "wrong" answer' (Vihvelin 2013: 114).

As is clear from the above, Vihvelin thinks that someone in an **Orwellian Choice** situation might possess free will even though such a situation appears to be a paradigmatic case of someone lacking all freedom. An agent such as Winston won't 'have a choice either about his overt actions or about his mental actions....but it remains true that [he] has a choice about what he *tries* or *begins* to do' (Vihvelin 2013: 113). Thus Vihvelin affirms that Winston has the *wide* ability to defy O'Brien, even if just momentarily; he can do this by *trying* to give the wrong answer. And Vihvelin says that Jones, the agent in a typical Frankfurt-style case, is no worse off than Winston (see the Introduction for a brief overview of Frankfurt-style cases; and see 3.4 for more detailed discussion). Jones is subject to potential intervention which curtails his range of freedom; he is only able to succeed in those actions (bodily and mental) that Black (the Frankfurtian intervener) wants him to succeed in. But like Winston, Jones is 'able to *try* to thwart his predetermined fate...Jones rerains alternatives [and] it is these alternatives, minimal as they are, that make it true that Jones retains his free will and his moral responsibility' (Vihvelin 2013: 114).

We can conclude, then, that Vihvelin does think that free will requires the wide ability to do otherwise. However, it is important to be clear that Vihvelin thinks this requirement will often be met in virtue of different kinds of wide ability. Sometimes, as in **Moorean Choice** situations, the requirement is met by an agent having the wide ability to decide otherwise and the wide ability to do otherwise. The agent would be able to successfully make a different decision and successfully perform a different bodily action on the basis of that decision. Interestingly, this requirement might be met even if the agent is unable to try to decide and unable to try to do otherwise. In **Shackled Choice** situations this requirement is met solely in virtue of the agent having the wide ability to decide otherwise; agents in such situations cannot perform different overt bodily actions, but they can perform many different mental actions. In an **Orwellian Choice** situation things are different again. Agents in such situations only have the wide ability to try to do something different; so although they cannot succeed even in performing a different mental action, they can – they have the wide ability to – at least try to do perform a different mental or bodily action.

Vihvelin's contention that free will does not require the wide ability to succeed in performing some alternative action is nothing new. Many of the earliest replies to Frankfurt's argument pointed out that those who think of free will as requiring (or simply being) the ability to do otherwise never took that to mean that the agent needed the ability to perform some other different action. Rather, the demand was that the agent be able to refrain from – or at least to begin to refrain from – the action performed (See Maria Alvarez (2009: 63–4) for some very useful discussion). What is distinctive in Vihvelin's account is that the agent might have the wide ability to do otherwise by sometimes possessing the wide ability to try, and other times without possessing the wide ability to try to do otherwise. There is much to question about this account. First and foremost is the wedge Vihvelin tries to drive between performing a different mental action and trying to do something: Vihvelin states that an agent in an **Orwellian Choice** situation *cannot* perform a different mental action but *can* try to do otherwise. But on many accounts, to try is to act. So if the agent can try to do otherwise, then the agent can perform a different mental action, namely, trying to do something different. Another worry concerns Vihvelin's characterisation of abilities as *dispositions to do something in response to the agent's trying to do it*. Given that characterisation, it's not clear what sense can be made of *the ability to try* – according to Vihvelin's account this would be understood as the *disposition to try in response to one's trying to try* – but this ability is crucial to Vihvelin's explanation of how it is that agents in **Orwellian Choice** situations and Frankfurt-style cases have free will, so it's crucial to her wider project. These issues will be taken up in the sections that follow and in the next chapter. I will begin in the next section by arguing that Vihvelin's account falls prey to the same problem that I deemed decisive against the simple conditional analysis of 'can': the **Transfer Problem**.

2.2. Vihvelin's account still plagued by the Transfer Problem

The primary problem I deemed decisive against the simple conditional analysis, the **Transfer Problem**, still affects Vihvelin's account. Here I will recast the problem in terms of trying, rather than choosing, in order to match Vihvelin's account of abilities. Recall from the previous chapter that Chisholm argued that the following three statements are consistent:

(**WC**) If Lehrer had tried to take a red candy, Lehrer would have taken a red candy
(**C1**) Lehrer would not have taken the red candy, had he not tried to take the red candy
(**C2**) Lehrer could not have tried to take the red candy

The conditional **WC** is the '**w**ould' **c**onditional used in the simple conditional analysis of the 'can' of free will. It is used to analyse the modal claim:

(**M**) Lehrer could have taken a red candy

The 'could' here is the 'could' of free will (i.e. a past indicative and not an incomplete subjunctive; cf. section 1.1.2). Chisholm argued that **C1** and **C2** together imply **~M**:

(**~M**) Lehrer could not have taken the red candy

And, as we saw, if Chisholm is right about this, then **WC** cannot be the correct analysis of **M** because it is consistent with two statements which imply the negation of **M**. To stand a chance of producing a correct analysis of **M** we would have to use the conjunction of **WC** and **~C2**. For example, we'd have to analyse 'Lehrer could have taken a red candy' as 'If Lehrer had tried to take a red candy, Lehrer would have taken a red candy, and Lehrer could have tried to take the candy'.

Vihvelin rejects this argument, saying that it fails not only against her final account, **LCA-PROP-Ability**, but also against the simple condition analysis (Vihvelin 2013: 203). Vihvelin grants that in many contexts where we utter counterfactuals like **WC** we do indeed assume that the corresponding claim like **C2** is false. That is, when we utter a sentence of the form 'If S had tried to X, S would have X-ed' we do typically assume that S could have tried to X (Vihvelin 2013: 200). Indeed, it may even be that in *most* such contexts we assume that S could have tried to X. But this, Vihvelin thinks, is irrelevant to the correct account of the agent's narrow ability to X. According to Vihvelin, there are two narrow abilities that we must clearly separate when considering this argument:

(**Try**) The narrow ability to try to X
(**Do**) The narrow ability to X

Vihvelin contends that an agent might possess an ability of either kind (*trying* or *doing*) without possessing the other – an agent might be able to X without being able to try to X, and vice versa. In the case described, Lehrer has the narrow ability to take the candy but not the narrow ability to try to take the candy. And we can easily imagine a case where someone has the narrow ability to try to take the candy but does not have the ability to take the candy – someone with extremely poor motor control skills, for example. Given that these abilities are independent in this way, Vihvelin maintains that it is illegitimate to demand of an account of the narrow ability to *do* X that it implies that the person also have the narrow ability to *try to do* X. And for this reason, it will also be illegitimate to demand of an account of the wide ability to do X that the person also have the wide ability to try to X (because possessing the wide ability is simply a matter of possessing a narrow ability and the associated opportunity). It is for these reasons that Vihvelin says that 'Chisholm's objection [the **Transfer Problem**] fails completely, not only against **LCA-PROP-Ability** and my claim that we have a narrow ability by having some intrinsic disposition, but also against the Simple Conditional Analysis' (Vihvelin 2013: 203).

The cases used to support the **Transfer Problem** typically involve an agent who suffers from some kind of pathology of the will, as with Lehrer's example where he is unable to try to take a red candy. As Vihvelin notes, this 'pathology doesn't function as the equivalent of a psychophysical paralysis preventing the person's choice or effective desire from causing the relevant movements of her body' (Vihvelin 2013: 203). That is, if the agent were, somehow, to try or choose, then the agent's trying or choosing would lead to the successful completion of the action tried. Vihvelin says that if this is the right way to think about these cases, 'then we should say that the person lacks one narrow ability – the ability to choose, on the basis of reasons, to take the candy ... – while retaining another narrow ability – the ability to take the candy ... as the upshot of choosing to do so' (Vihvelin 2013: 203). By 'upshot' here, Vihvelin means 'as an effect of choosing'. In other words, an agent has the ability to X because if the agent were to try to X, the agent's trying to X would cause the agent's X-ing. But this does not require that the agent be able to try to X. In the **Red Candy** example, Lehrer (assuming he is in the vicinity of a candy) will, on Vihvelin's view, have the wide ability to take the candy but lack the wide ability to try to take the candy. Vihvelin thinks that possessing this ability will suffice for having free will. And that is why she thinks Chisholm's complaint is misguided: the ability to try to X comes apart from the ability to do X, and possessing either (even without possessing the other), suffices for possessing free will.

Vihvelin, I think, is right about one thing: it is possible for an agent to have either the narrow or the wide ability to X without also having the narrow or the wide ability to *try to* X, and vice versa. For example, according to Vihvelin's accounts, Lehrer has the narrow ability to take the candy (because he would take it, if he were to try), but he doesn't have the narrow ability to try to take the candy. And someone with severely impaired motor control might well have the narrow ability to try to take the candy, but not the narrow ability to actually take it. Still, Vihvelin's supposed refutation of the **Transfer Problem** does not succeed. This is because it presupposes that her classification of abilities – into skills, narrow abilities and wide abilities – is exhaustive. If this classification of abilities were exhaustive, then upon learning that Lehrer had the wide ability to take the candy we would be forced to conclude that Lehrer had free will with respect to whether he took the candy – this follows on the fairly plausible assumption that wide abilities are, of the three, those most relevant to a person's free will. To put it another way: to have free will with respect to taking the candy, we want to consider Lehrer's 'taking the candy abilities', and out of the skill, the narrow ability and the wide ability to take the candy, it will be the wide ability which matters most.

However, Vihvelin's classification of abilities is not exhaustive. There is, it seems, a sense of 'able' that includes the idea that the agent possesses everything that is needed in order to perform some action, bar those things that the agent can acquire by acting. In other words, and to use Vihvelin's terminology, there is a sense of 'able' which expresses the idea that someone possesses both the wide ability *to try to X* and also the wide ability *to X*. And it's this sense of 'able', Chisholm and Lehrer would say, that is relevant to free will. The thought driving Chisholm's objection is that it must be possible, in some fairly strong sense, for the agent to X. If trying to X is the only route to the agent's X-ing then it must be possible for the agent to try to X. Vihvelin's diagnosis of Chisholm's argument is thus mistaken. It is not, as Vihvelin argues, that Chisholm thinks lacking the narrow ability to choose automatically means lacking the narrow ability to act on the basis of the choice made. Nor does he think anything similar is true of the wide ability to choose and the wide ability to act on a choice made. Remember, Chisholm is not providing an account of intrinsic abilities – as Vihvelin is, with her narrow abilities – and nor has he provided a classification of the different kinds of ability. Chisholm is discussing the sense of 'able' relevant to free will directly and his point is that if an agent is unable to try – and if the agent must try in order to succeed – then the agent is unable in the sense relevant to free will.

Does Vihvelin's own account fall prey to Chisholm's objection? The answer is complicated because Vihvelin does not think there is any one kind of ability which is needed for free will. Wide abilities are always involved. As previously discussed, however, there is variation in which wide abilities are required. The free will that 'we ordinarily think we have' involves possessing the wide ability to deliberate, decide and do something other than what we in fact did. And it's interesting to note that on Vihvelin's account someone might possess all of these abilities without yet being able to *try* to deliberate, decide or do something different. It might be questioned whether it is possible to try to deliberate or to try to decide. On Vihvelin's account, however, *trying* consists in one of the agent's intentions causing (the beginnings of) some behavioural process that the agent believes (perhaps wrongly) will bring about, or at least move the agent closer to, the intended goal (Vihvelin 2013: 175). So for Vihvelin it *does* make sense to talk about abilities to try [i.e. begin] to deliberate, and abilities to try to decide. Moreover, on Vihvelin's account it's possible to have the ability to deliberate – the ability to succeed in deliberating if one tries – without yet being able to try to deliberate. As Vihvelin understands it, then, having the free will 'we ordinarily think we have' does not involve having the ability to try or begin any of those things which we have the ability to do. As a result, Vihvelin's account of the free will 'we ordinarily think we have' will indeed fall prey to the **Transfer Problem**.

The complication comes when we consider those agents who've had some of their free will curtailed. As we saw, Vihvelin said that someone who is in an **Orwellian Choice** situation, and who is thus subject to such physical and mental pressures that they cannot perform any different bodily or mental actions, does have the wide ability to try to do something different. Vihvelin's account of such agents will therefore not succumb to the **Transfer Problem** because these agents have the ability to try to do otherwise. Ironically then, on Vihvelin's construal, those people who are in what appears to be a rather dire situation – the **Orwellian Choice** situation – actually turn out to have a more robust kind of free will than those who are in an **Moorean Choice** (i.e. everyday choice) situation; the former have all the abilities they need (namely, the ability to try) to attempt to do something else, whereas the latter might well be lacking a crucial necessary condition (namely, the ability to try) for doing what it is they have the ability to do.

It is worth noting that Moore himself recognised the need to affirm that trying (or choosing) was a possibility. Towards the end of the chapter on free will in his *Ethics* he imagines an objector pushing the following point:

> Granted that we often [would] have acted differently, if we had chosen differently; yet it is not true that we have Free Will, unless it is also often true in such cases that we could have chosen differently (Moore 1912: 218).

Moore seems to agree with this point: he offers two different responses, both of which involve affirming that there is indeed a sense in which the agent could have chosen differently. Moore's first line of response appears to involve applying his conditional analysis of 'can' to the agent's being able to choose:

> If by saying that we could have done, what we did not do, we often mean merely that we [would] have done it, if we had chosen to do it, then obviously, by saying that we could have chosen to do it, we may mean merely that we [would] have so chosen, if we had chosen to make the choice (Moore 1912: 218–9).

And Moore goes on to say that 'there is no doubt [that] it is often true that we [would] have chosen to do a particular thing *if* we had chosen to make the choice' (Moore 1912: 219). These comments suggest that Moore means to propose something like the following:

> 'S could have chosen to A' is true if and only if S would have chosen to A, if S had chosen to make the choice.

All this does, however, is multiply problems for Moore. First, there is the threat of a regress, for now we have to ask whether the agent could have *chosen to make the choice*. If Moore gives a positive answer, then he owes us an analysis of the agent's being able to choose to make the choice. And clearly, he cannot just keep applying the conditional analysis. Second, it's not clear whether there is any way to understand the agent's *choosing to make the choice* which helps Moore. An example from van Inwagen (1983: 116–7) helps here. Suppose that Peter faces a choice between drinking claret at dinner and drinking burgundy at dinner. We can understand what it would be for Peter to choose, at t1, that he will make the choice between claret and burgundy at, say, t2. Perhaps Peter wants to see how he feels later on, or is too busy at t1 to think properly about it, and so on. Whatever the reason, we can make sense of someone who at some point decides that they will make some other decision at some later time. But if this is how Moore intends the 'choosing to make the choice' to be read, then he faces a further two problems. First, this reading threatens the truth of the conditional 'S would have chosen A, if S had chosen to make the choice' because the antecedent now refers to a prior choice of the agent's which consists in deciding to decide the matter later on. But if the agent is merely deciding to settle the matter later on, then at the very least it would be highly controversial to affirm the consequent, namely, that S would have chosen A. That is, if Peter decides (at t1) that he will make the decision between claret and burgundy later on, it is at best highly controversial whether (at t1) there is any fact of the matter concerning which drink Peter will choose. Second, this reading now appeals to a prior decision by the agent: the deciding to make the decision later on. And so now we have to ask whether and in what sense the agent could have made this prior decision.

A different way of reading 'S would have chosen to make the choice' has it that S chooses at t1 to make a decision to pick a particular option at some later point. In other words, on this second reading 'choosing to make the choice' is not a matter of the agent choosing at t1 to make the decision at t2, but rather it's a matter of choosing at t1 to decide on a particular course of action at t2. To use the previous example once more, the suggestion would be that Peter decides at t1 that he will make a decision to choose claret at t2. I concur with van Inwagen that little sense can be made of this (van Inwagen 1983: 116). A decision is essentially an action which resolves some uncertainty concerning what the agent is to do; a decision leaves the agent with an intention to act in some particular way. So an agent cannot decide now to make a future decision in some particular way. For this implies both that the agent has settled the matter about how to act at that future time and also that how the agent is going to act is unsettled (because there is still a decision to make about it). So this second reading of 'choosing to make the choice' is also of no help to Moore.

Perhaps an awareness of these problems led Moore to propose his second line of response, which was to claim that there is a different sense in which the agent could have chosen otherwise, namely, that *it was possible that* the agent choose otherwise. What sense of possibility is Moore invoking? He's very clear:

> This sense arises from the fact that in such cases we can *hardly ever know for certain* beforehand, which choice we actually [will] make; and one of the commonest senses of the word 'possible' is that in which we call an event 'possible' when no man can know for certain that it will not happen (Moore 1912: 219; emphasis in original).

In other words, Moore's suggestion here is that we can affirm that the agent could have chosen otherwise because, as will surely be the case in most choice situations, *for all anyone knows* the agent might have chosen otherwise: i.e. it was epistemically possible that the agent choose otherwise. Moore is correct to suggest that this claim would be compatible with the thesis of causal determinism. But he's incorrect to suggest that affirming that it was epistemically possible that the agent chose otherwise goes some way to addressing Chisholm's worry. Chisholm's argument is that if there is something the agent needs to be able to do in order to X, but which the agent cannot do (all things considered), then the agent cannot (all things considered) X. It might well be true that it was possible that the agent choose to X inasmuch as the factors which precluded the making of that very choice were unknown to anyone, but that is irrelevant to Chisholm's argument. That is not to say, of course, that the agent's doxastic and epistemic state is irrelevant to the control they can exert. I will argue in chapter 5 that the control exerted by an agent depends in significant ways on the beliefs that the agent has; but the doxastic or epistemic openness required is not instead of the requirements suggested by Chisholm's argument but in addition to them. The primary purpose of bringing Moore back in at this point, however, is just to note that at least Moore recognised the need to affirm that the agent could have tried or chosen otherwise. Vihvelin, on the other hand, denies this and for that reason her account fares worse than Moore's.

2.3. Vihvelin on the nature of abilities

2.3.1. Intrinsic abilities are not individuated by action type alone

In this section I will argue that there is a problem lurking for Vihvelin's account which concerns the individuation of dispositions and abilities. Aside from the sense of 'ability' which refers to skills, Vihvelin recognises two senses of 'ability': narrow ability and wide ability. Narrow abilities are intrinsic, wide abilities are extrinsic. Thus, ignoring the skill sense of 'ability' for the time being, when we say that Olivia is able to leave her office there are (according to Vihvelin) just two properties that we might be attributing to Olivia: the narrow ability to leave her office or the wide ability to leave her office. In what follows I will argue that this is a mistake. Even restricting our focus to intrinsic abilities, a phrase like 'is able to leave her office' does not succeed in picking out a unique ability. Given that Vihvelin treats abilities as dispositions, I will begin by demonstrating the point for dispositions.

Consider then the boiling behaviour of water. Water has the disposition to boil at 100 degrees C. But it exhibits this behaviour only if the pressure is around 1 atmosphere. At a pressure of 2 atmospheres water will boil at roughly 119 degrees C. Similar things can be said of solubility. There is a straightforward sense of the term 'soluble' according to which salt is soluble while nail polish is not. But there are circumstances under which nail polish will dissolve: when placed in water at high temperatures and pressures, or when placed in acetone at standard temperature and pressures (Prior 1985: 5–6). So 'soluble' as used in ordinary language must mean more than simply 'is disposed to dissolve when placed in water'. We need to clarify the disposition being expressed with something like 'is disposed to dissolve when placed in water at standard temperature and pressure'. If we didn't add this extra information, if 'soluble' simply meant 'is disposed to dissolve when placed in water' or 'is disposed to dissolve under some circumstances or other', then the term would not be able to play the classificatory role it does. Elizabeth Prior summarises this point nicely: 'dispositional predicates would lose [the power to classify objects] if our criterion for ascribing disposition D to an item were simply that that item would manifest that disposition under *some* set of conditions' (Prior 1985: 6).

When we add this detail we flesh out the characterisation of an *intrinsic disposition*. It's not that, for example, some water currently in an environment where the pressure is 1 atmosphere has the disposition to boil at 100 degrees C only because it is in that environment, and that when moved to an environment with a pressure of 2 atmospheres it would lose one disposition and gain another. Nor is it the case that some nail polish stood on a dressing table has no disposition to dissolve, but that if it were placed in some acetone it would there and then gain the disposition to dissolve and also start manifesting that disposition.

Rather, water has the disposition to boil at 100 degrees C when the pressure is 1 atmosphere *regardless of the circumstances it is currently in*. That is something true of water – a property we can ascribe to it – without ever knowing what circumstances the water is in. Similarly, nail polish has the disposition to dissolve in acetone at standard temperature and pressure, as well as, e.g., the disposition to dissolve in water at very high temperatures and pressures, even if it is not currently in either of those situations (Prior 1985: 6–7). Some nail polish would still have those dispositions even if it were locked in a safe in the middle of a desert.

Yet because we need to include such information simply to understand which property we are ascribing – there is, after all, a real distinction between the dissolving behaviour of salt and nail polish even though both dissolve in water under some conditions or other – the information added must be part of the characterisation or definition of the disposition in question. In addition to the stimulus conditions and the manifestation type, then, some set of circumstances are also needed in order to properly characterise the disposition. I will call this set of circumstances the *definitional circumstances*.

Here is an example which illustrates this point for abilities: Ann plans to take her elderly father, Abe, out for coffee and, having not visited in a while, she asks her brother whether their father can walk or whether he'll need the wheelchair (suppose that Abe is getting frail, but likes to walk whenever possible). Knowing that the route from her father's bungalow to the car is less than 50m and across a flat driveway, and that the route from the car park to the coffee shop will be similar, Ann is asking whether Abe has the ability to walk very short distances across flat surfaces. She's not asking whether Abe has the ability to walk up inclines of 10%, nor is she asking whether he can walk for 40 minutes straight or in heat of 40 degrees C. Furthermore, Ann is not asking whether Abe is able to walk in 'ordinary conditions' – whatever those are supposed to be. She's interested solely in whether Abe has a power or ability to do something in a relatively narrow range of situations. Ann asks about this ability because of the interests she has and different interests might lead to different questions: Ann's brother might query whether his dad can walk a few meters down a 10% slope because he has a very steep driveway.

Once again, both Ann and her brother are asking about Abe's intrinsic abilities. The move from talking about 'the ability to walk' to 'the ability to walk very short distances across flat surfaces' need not mean a move from talking about intrinsic abilities to talking about extrinsic abilities. The details added to the description of the ability concern one potential external environment that Abe might find himself in – but they *do not* play the role of specifying those circumstances in which the ability is *possessed*. Ann is *not* asking about whether Abe *would have* the ability to walk (*simpliciter*) *if he were* near a flat surface. Rather, she is asking about whether Abe has an ability, the full characterisation of which is *the ability to walk very short distances across flat surfaces*. The detail that we've added here becomes part of what I called above the *definitional circumstances* – that set of circumstances which, together with the stimulus conditions and the manifestation type, characterises or defines the ability.

This thought – that we're still talking about intrinsic abilities – can be supported by highlighting the difference between definitional circumstances and opportunities: if an agent has, at t, an intrinsic ability to A, nothing is said one way or the other about whether the agent is, at t, *in* circumstances which are of the same type as the ability's definitional circumstances. For example, I may have (at t) the ability to walk but might not be (at t) on a hard surface – the latter arguably being a precondition for all successful walking. What's being said, roughly, is that *were* the agent in such circumstances, a certain action *would be* possible. By contrast, if I have an opportunity to walk, then *I am in* a set of circumstances which match the type of the definitional circumstances for some corresponding ability to walk.[7] Definitional circumstances concern a set of circumstances (the set in which it is possible for the agent to perform an action) which *may or may not* be actual; opportunities concerns *the actual* circumstances that the agent is in.

A further way to underscore the point is as follows. It might be that Abe *is able* to walk short distances across flat surfaces and yet *unable* to walk down steep slopes. Alternatively, it might be that Abe is able to walk short distances across flat surfaces and also able to walk down steep slopes. It seems that either of these scenarios might be case even when, for example, Abe is locked in a room and so not in a position, there and then, to exercise either of these abilities. Still, there would be a real difference between the scenarios: Abe would be more able in the latter. But if Abe might possess those abilities when he is locked in a room then they must be intrinsic abilities and the information which we've added to the description of the ability must be characterising the intrinsic ability rather than describing Abe's extrinsic circumstances.

The key claim of this section, then, is that the stimulus conditions and manifestation type of intrinsic dispositions and abilities – at least as they are typically specified in both ordinary language and philosophical discussions – are not sufficient to pick out a unique property. In addition, some set of circumstances is needed in order to complete the characterisation or definition of the property. But as was suggested above, and as was evident from the various examples, different sets of circumstances may be combined with the same stimulus and manifestation pair to produce different dispositions or ability properties. To put it another way: the definitional circumstances are not a function of the stimulus conditions and/or the manifestation. This gives us, if you like, a spectrum of dispositions and abilities for each different manifestation type: there are many dispositions to dissolve and many abilities to walk. Vihvelin, by not fully appreciating the role that

[7] Perhaps the idea of opportunity is wider than this and includes my being in circumstances where I could easily get into a position to exercise an ability. Whether or not this is so, and how exactly such a qualification could be spelled out, is a non-trivial matter, but it does not affect the substance of the point being made.

circumstances play in defining dispositions and abilities, misses this. I will show in the following sections and the next chapter how this causes a problem for her account.

It might be thought that Manley and Wasserman's account of dispositions has a ready explanation for the examples cited above, and can therefore avoid positing anything akin to definitional circumstances. This is because on Manley and Wasserman's account, dispositions are defined in part by a modal force parameter. If the examples given can be explained in terms of one object's being *more disposed* than another, or one agent's being *more able* than another, then my argument will be ineffective against their account. For example, salt and nail polish clearly differ in their behaviour when it comes to dissolving in water. But it might be suggested that this difference can be explained purely in terms of salt being *more water soluble* than nail polish. That is, it might be suggested that in quantifying over all the possible 'being placed in water' cases (which share our laws) we will find that salt dissolves in more of them than nail polish. This could then be used to explain why sometimes we call salt and not nail polish soluble: in such contexts we are setting the bar high, as it were, and only calling substances which dissolve in a relatively large number of the 'being place in water' cases soluble.

Manley and Wasserman's account does, I think, have a way of explaining some of the examples I presented. As the preceding discussion has just illustrated, their account seems to provide a good explanation of the differences between salt and nail polish with respect to water-solubility. Things are less clear with other examples, however. Consider once more the Ann and Abe example. On Manley and Wasserman's view, there is just one ability to walk that is possessed by different agents to different degrees. To explain the Ann and Abe example, Manley and Wasserman would have to say that Ann and her brother are both asking about the ability to walk but that each is asking about whether that ability is possessed to a different degree. This might seem initially plausible: if Abe possesses the ability Ann's brother is interested in (the ability to walk 10m down a very steep slope) then it might be thought that Abe will also possess the ability Ann is interested in (the ability to walk a short distance along a flat surface). Walking down steep hills is, intuitively at least, 'more difficult' than walking along flat surfaces, so if Abe has the former ability he will have the latter. And that would enable a Manley and Wasserman's account to explain the data: their account quantifies over all possible walking scenarios, including both flat surfaces and down steep slopes. In asking about walking along flat surfaces, Ann is setting the threshold fairly low whereas her brother is setting it a bit higher.

However, given the embodied nature of agency, it is by no means certain that someone who can perform what is intuitively a more difficult task can also perform an easier one. It might be true in many cases, but it is not necessarily so. For example, frail Abe might have enough balance and control to be able to learn how to take advantage of the downward slope, such that he can exploit gravity and so successfully traverse his son's driveway. At the same time, his heart and muscles might not be strong enough to propel him more than 10 metres across a flat surface. Here, Abe would have the ability his son is interested in but not the one his daughter Ann is interested in. So while Manley and Wasserman's account has some success in handling some of the problem cases, it is by no means clear that they can explain all such cases. Moreover, any account involving quantification over all possible stimulus cases is still subject to the epistemological worry that Vihvelin raised, namely, that we only have knowledge about the tiniest proportion of such cases.

Before closing this section I want to be clear about one thing the view being proposed is not committed to. The line of argumentation above bears some resemblance to a view recently proposed by Justin Fisher (2013) to the effect that the canonical form of dispositions should involve three parameters and not two. That is, the canonical form of a statement ascribing a disposition should be as follows, where (following Fisher's notation) braces mark the positions of the parameters:

> Object O is disposed to {M} when {S} in {C}

This contrasts with what is ordinarily taken to be the canonical form of such statements:

> Object O is disposed to {M} when {S}

I am sympathetic to thinking of the definitional circumstances as a third parameter in the way Fisher describes. However, accepting the argument above does not commit us to this way of thinking. The key claim I'm making is that the stimulus conditions and the manifestation type *as typically specified* (in both ordinary language and philosophical discussions) are not sufficient to pick out a single property. The event types typically cited (e.g. breaking, dropping, dissolving) are just too broad. This is compatible with saying that the extra information needed – what I have called the *definitional circumstances* – is best thought of as part of the stimulus conditions or, indeed, the manifestation. As Fisher (2013: 450) himself says, perhaps the canonical form of such statements involves just two parameters and the definitional circumstances are best incorporated into either stimulus conditions or the manifestation:

> Object O is disposed to {M-in-C} when {S}
> Object O is disposed to {M} when {S-in-C}

If this were right, then the precisified stimulus and manifestation would be enough to fully characterise the disposition or ability. The argument presented above does not require us to take a stand on this issue. It is also worth mentioning that the motivation Fisher presents for his view is closely tied to his desire to give an analysis of dispositions in terms of counterfactuals: with an analysis of dispositions in terms of counterfactuals as a desideratum, Fisher thinks that it is necessary to introduce something like definitional circumstances in order to solve the problem of masks. The argument I've presented above is intended to be prior to any worries concerning the problem of masks.

It will be useful to introduce a convention to indicate when a set of circumstances forms part of the definition of a disposition or ability. I will follow Bird (1998: 232) and Ann Whittle (2010) and use hyphens to indicate this. For example, the salt/nail polish example could be summarised as follows: salt has, and nail polish lacks, the-disposition-to-dissolve-in-water-at-standard-temperature-and-pressure, while both salt and nail polish have the-disposition-to-dissolve-in-water-at-very-high-temperatures-and-pressures. This convention is a little ugly, and I will attempt to use it sparingly, but it is useful when emphasising the difference between definitional circumstances and opportunities.

2.3.2. Vihvelin's account of narrow abilities and the getting specific strategy

In this section I will argue that Vihvelin's account struggles to accommodate the points made above concerning the individuation of abilities. Recall that Vihvelin's final account of narrow abilities ran as follows:

> **LCA-PROP-Ability**
> S has the narrow ability at time t to do R in response to the stimulus of S's trying to do R **if and only if**, for some intrinsic property B that S has at t, and for some time t' after t, **if** S were in a test-case at t and S tried to do R and S retained property B until time t', **then** in a suitable proportion of these cases, S's trying to do R and S's having of B would be an S-complete cause of S's doing R (Vihvelin 2013: 187).

In **LCA-PROP-Ability**, 'R' stands for the manifestation of the disposition which will be the action type of the ability (Vihvelin calls it the **r**esponse) while the stimulus conditions have the form of *S's trying to do R*. The abilities ascribed by this account will therefore have the form *the narrow ability to do R in response to the agent's trying to do R*. Such properties appear to be individuated by the action type alone – there is no mention of any circumstances resembling the notion of definitional circumstances outlined above.

Things are complicated somewhat, however, by Vihvelin's appeal to the notion of a test-case. Recall from section 2.1.3 that test-cases are those possible scenarios which are *relevant* or which *matter* with respect to the testing of whether an ability is possessed. Vihvelin's **LCA-PROP-Ability** says that an agent has the ability to A only if that agent A-s in a suitable proportion *of valid test-cases for A*. In other words, whether or not an agent has an ability to A depends on what happens in a certain, clearly defined set of circumstances. This might suggest that Vihvelin's account defines abilities not just in terms of the action type but also in terms of a set of circumstances, perhaps akin to what I have called the definitional circumstances.

Whether or not this is so depends on whether the test-cases are a function of the action type, R. That is, in order to accommodate the points outlined in 2.3.1 it is not enough to hold that there is *a* set of circumstances which contributes something to the definition of the ability; one needs to hold that a given manifestation type (here, R) may be paired with different sets of definitional circumstances (thus producing different abilities). But one might hold, as Lewis (1997: 153) did, that a property such as *is disposed to cause death when ingested* is only a rough characterisation of a disposition, but that there is just one way it is to be filled out, as determined by the getting specific strategy (see 2.1.3 for full details). In other words, the getting specific strategy is applied to a stimulus/manifestation pair and yields a fully characterised disposition, there being only one way to apply the strategy for each stimulus/manifestation pair.

This assumption – that the getting specific strategy may be applied in only one way – appears to be a common assumption of those who entertain its use (Manley & Wasserman) and those who explicitly endorse it (Vihvelin). This is not surprising: the idea follows naturally from the purpose for which the strategy was originally introduced by Lewis, namely, to solve the problem of masks. Recall that a mask is a factor which, if it occurred alongside the stimulus conditions, would interfere with the manifestation of a disposition. Masks are problematic for those who hope to employ some kind of conditional analysis or account of dispositions, as Lewis did and Vihvelin does. And they are easy to find when a disposition is characterised in terms of broad stimulus and manifestation events (e.g. 'is disposed to cause death when ingested'). This is why Lewis (1997: 153) calls such statements 'rough characterisations' of dispositions. The idea of the getting specific strategy is that by adding detail the counter-examples will be ruled out; for example, we take a 'rough' dispositional statement like 'is disposed to cause death when ingested', apply the getting specific strategy, and end up with a fully specified and mask-safe definition of a disposition. Given that project, the question of whether there might be more than one way to apply the strategy never comes up.

Vihvelin's use of the getting specific strategy matches this pattern and this means that it does not recognise the full spectrum of ability properties. All is not lost for Vihvelin's account of abilities: it can easily be amended to solve this problem. What I will argue in the next chapter, however, is that if we make this amendment, then Vihvelin's account of the ability to do otherwise, the compatibilist account of free will she builds on top of it, and her treatment of Frankfurt-style cases are all undermined.

In the remainder of this section I will argue against the standalone use of the getting specific strategy. That is, I want to suggest that the getting specific strategy simply does not work unless a rough idea of the definitional circumstances is already possessed. The getting specific strategy, in other words, may be used to refine a set of definitional circumstances but it cannot be used to arrive at some set of definitional circumstances (out of nothing, as it were).

To see this, consider the following example: suppose you have seen Manuel memorise the order of a shuffled pack of cards on a number of occasions. You also know, however, that Manuel fails at this task whenever there is too much ambient noise. Now consider this question: does Manuel have the ability to memorise the order of a shuffled pack of cards? On Vihvelin's account, the answer to this depends on whether Manuel succeeds in memorising a pack of cards in a 'suitable proportion' of relevant test-cases. But what counts as a relevant test-case? To answer *this* question Vihvelin will apply the getting specific strategy. And much hangs on the result: if the background noise counts as a mask then it should – on Vihvelin's view – be excluded from the definition of the ability. As a result Manuel will have the ability to memorise a pack of cards despite his failure in the ambient noise cases. If lots of background noise doesn't count as a mask then whether someone possesses this ability will depend in part on what happens in such cases. Manuel's failure in these cases will therefore potentially undermine his possession of the ability. But it's not at all clear that the getting specific strategy can answer the question whether background noise counts as a mask.

When Lewis hinted at the getting specific strategy he used the example of *being poisonous*. One difference between the property *being poisonous* and the ability being ascribed in the Manuel example is that the former is an everyday concept for which we have (at least fairly) clear intuitions about when it applies. My contention is that our understanding of the term 'poisonous' includes some understanding of what I have called definitional conditions. It is knowledge of such conditions that allows us to proceed with the application of the getting specific strategy.

The memorisation example is useful because its unusualness precludes us having any firm intuition concerning the two scenarios. And this helps make it clear that the correct answer as to whether background noise should count as a mask in the Manuel example depends not on the result given by the getting specific strategy – whatever that might be – but on what we mean to say in ascribing the ability to Manuel. Here are two ways the story could be filled out which illustrate this point. Suppose that Manuel is hoping to use his memory skills to his advantage in a casino. If he is to succeed in this endeavour, he will need to have the ability-to-memorise-a-pack-of-cards-whether-or-not-there-is-background-noise. If this is the ability in question then the presence of background noise will not count as a mask; rather, situations with much background noise are paradigm test conditions. If Manuel invariably fails to memorise a pack of cards when there is background noise, it's not that he has the ability mentioned but that it is masked by the background noise. Rather, he simply fails to have the ability: he fails to have the ability that would matter in the casino. In Vihvelin's terminology, cases including background noise count as valid test-cases for the ability in question and so failing to exhibit the behaviour in such cases will count against Manuel's possessing that ability.

On the other hand, suppose that Manuel is interested in entering a memory competition. In this case what's important is whether he's able-to-memorise-a-pack-of-cards-with-no-background-noise (because these are the only conditions under which he'll be tested). If during a practice in front of friends, a loud noise disturbs him and he fails to memorise the pack, his friends would miss the mark if they advised against entering on the grounds that Manuel didn't have the requisite ability. Manuel would rightly point out that the loud noise was something which interfered with (i.e. masked) his ability and in making this claim he'd be referring to his ability-to-memorise-a-pack-of-cards-with-no-background-noise.

To re-iterate a point made earlier then, each of these abilities is an intrinsic ability. The circumstances in question *complete the definition* of the ability property. They are not simply those circumstances in which the ability is *possessed*. Nor are they a description of some *opportunity* that Manuel has. To say Manuel has the-ability-to-memorise-a-pack-of-cards-when-there-is-no-background-noise is not to say that Manuel has the ability-to-memorise-a-pack-of-cards and he has an opportunity to exercise that ability. In attributing the former ability nothing is said one way or the other about the circumstances Manuel is actually in. Moreover, Manuel's possession of each of the two abilities mentioned is independent of his interests. Of course, which ability he (or his friends) pick out with the phrase 'the ability to memorise a pack of cards' will depend on their interests, but Manuel's possession of these abilities is not interest-relative.

As mentioned, Vihvelin's account could be amended to cope with the above points easily enough. We could, for example, introduce a third parameter (a set of circumstances) into the explanandum alongside the action type and then have the getting specific strategy be a function of these circumstances and the action type. The problem for Vihvelin is not that her account cannot be so amended, but rather that amending it in this way undermines her case for saying that agents in Frankfurt-style cases are able to do otherwise, which is a large part of her case for saying that the thesis of causal determinism doesn't undermine the ability to do otherwise. I will establish these points in the next chapter.

Chapter 3 – Abilities and circumstances[8]

3.1. A recap

I will begin this chapter by drawing together some of the key points made in the previous chapter concerning the nature of dispositions and abilities. There I argued that the event types typically given when specifying an ability are too broad to individuate the ability; a set of *definitional circumstances* needs to be added to the definition of the ability. The definitional circumstances could be glossed as 'the circumstances to which the ability applies'; that is, they are the circumstances in which, given the stimulus conditions, the manifestation behaviour (for abilities, the performance of the action) is to be expected. In addition, I argued that neither dispositions nor abilities require that when the stimulus conditions obtain the manifestation *always* occurs. To accommodate this I endorsed the view that the modal force – intuitively, the 'strength' of the ability or disposition – need not be any kind of restricted necessity (see 2.1.3 for full details). Moreover, I suggested that the modal force too should be considered part of the definition of the disposition or ability because objects might have dispositions or abilities of different strengths. The brick is not very breakable but the glass is: the brick's disposition to break has a low modal force, the glass's disposition to break has a high modal force (again, see 2.1.3). The same is true of abilities: two people might have the ability-to-cycle-in-heavy-traffic but to different degrees. Finally, I noted that some dispositions (and perhaps some abilities) do not seem to have any stimulus conditions at all (e.g. the disposition of a radioactive particle to decay; loquaciousness). So the stimulus conditions are best seen as optional. This gives us the following list of things which play some role in individuating dispositions and abilities:

Definitional circumstances (a background against which the manifestation is to be expected)
Manifestation (the behaviour which ensues; in the case of abilities, the action type)
Modal force (the 'strength' of the disposition or ability)
[Optional] A set of stimulus or triggering conditions

There is one caveat to be noted: strictly speaking the idea isn't that you *cannot* define an ability using very coarse event types for the stimulus, manifestation and definitional circumstances. Rather, the point is that such a property will not be very informative. Such a property says that in some proportion – settled by the modal force of this particular disposition – of the range of possible scenarios demarcated by the definitional circumstances (and further constrained by the stimulus conditions, if there are any), the behaviour ensues. But if the definitional circumstances (and the stimulus conditions, if there are any) are too broad, then the range of possible scenarios under consideration will be so broad as to render the claim being made fairly uninformative. Consider an example from Vihvelin (2013: 185). Matches have the disposition to light when struck. If we let

[8] Portions of chapters 2 and 3 were previously published in Kittle (2015). Thanks go to Springer Netherlands for allowing inclusion here.

the definitional circumstances be the entire range of nomologically possible striking cases, as in Manley and Wasserman's account, then to say that a match has the disposition to light when struck with modal force F is to say that in proportion F of all such striking cases, the match will light. But the striking cases being considered are so diverse – they include cases where the match is struck in the middle of a tornado, on the top of a mountain, on the moons of far off planets, and so on – that (assuming the match does have the said disposition) we still can't say very much about the circumstances in which it would light. The possible cases where the match does light when struck might be very homogenous or they might be spread out very thinly across 'modal space' (or something in between). And any object which lights in enough cases – no matter how diverse – would count as having the disposition to light when struck. But that means we could not use the disposition to light when struck (as it is currently being understood) to classify objects. To be sure, the matches we encounter light when struck in many typical, everyday circumstances. But it might be that there are many possible circumstances where a pen lights when struck. If so, it would count as having the same disposition as the match in virtue of those possible cases where it lights. And this makes dispositions and abilities – when they are defined by definitional circumstances and stimulus conditions that are very broad – useless when it comes to classifying objects, which is one of the main uses of such ascriptions.

This section considers further the nature of abilities when they are individuated in the way suggested. Although Manley and Wasserman ascribe properties that are not very informative (for the reason given immediately above), their understanding of those properties is a huge advance over conditional accounts which employ the standard Lewisian semantics for counterfactuals. This is for two reasons. First, the latter theorists are committed to saying that the modal claim expressed by ascriptions of dispositions and abilities is a form of restricted necessity. Such analyses are therefore incapable of accommodating exceptions – cases where the stimulus conditions obtain but the manifestation does not occur. This is why many of those who want to retain the conditional analysis – e.g. Lars Gundersen (2002), Sungho Choi (2008) – do so by appealing to non-standard accounts of the semantics of counterfactuals. Second, the use of the Lewisian counterfactual threatens to make the *possession* of all dispositions an extrinsic matter. This is because on Lewis's semantics the truth of a counterfactual depends on similarity between possible worlds; but that similarity will invariably involve how similar the worlds are in various 'local matters of fact'. Most of the pertinent 'local matters of fact', however, will be ones that are extrinsic to the object or person possessing the disposition or ability. Manley and Wasserman dub this the 'accidental closeness' worry (Manley and Wasserman 2008: 70). It means, for example, that if a sturdy non-fragile brick is precariously perched on a window ledge 60 metres above the ground then the counterfactual 'if the brick were pushed (or dropped, or struck, etc), it would break' will come out true – because the closest world where it's pushed or dropped will involve it falling 60 metres to the ground – despite the brick not being fragile.

Manley and Wasserman's direct quantification over possibilities addresses these concerns and is, therefore, a big improvement. The scheme I am proposing retains this feature, but it allows for variation in the set of possibilities quantified over. In section 3.2 I will explore further this understanding of abilities; in particular, I will discuss the relationship between abilities on the spectrum I envisage (as introduced in 2.3.1). The aim is to answer the following question: when is an agent able to do otherwise in the sense required by free will? This question will be directly addressed in 3.3 using the groundwork laid in 3.2. This chapter will close by considering how Vihvelin's account of abilities, discussed at length in the last chapter, is affected by the points made in 3.2 and 3.3.

3.2. The nature of abilities and how they relate to one another

I have argued that dispositions and abilities are in part defined by a set of *definitional circumstances*. These are circumstances which are intended to augment typical characterisations of the stimulus and manifestation type. But although different definitional circumstances can be paired with the same manifestation type, the latter puts some constraints on the former. This is because in order for something to count as an instance of a given manifestation type, some minimal preconditions typically have to be met. Consider the ability to walk. There are certain situations where walking is impossible because the circumstances don't allow it. For example, and very roughly, walking requires a surface hard enough to step on and enough downwards force to stop the agent floating upwards. If we ascribe to an agent the ability to walk we say nothing about what the agent is able to do when these kinds of minimal conditions are not met. I am not here thinking about cases of finking or masking (see 2.1.3). Rather, I am getting at the idea that in some circumstances the ability simply 'fails to apply' – it isn't applicable to the person's circumstances in the first place. The same is true of dispositions. Salt has the disposition to dissolve in water. But the solubility of a lump of salt in a desert is not being masked; rather, the minimal conditions needed in order for the dissolution behaviour to be a 'live possibility' are simply not met.

I will call the minimal conditions that need to be met in order for an ability to be exercised the *minimal action realisation conditions*; and I will call the minimal conditions needed for the manifestation of some disposition the *minimal manifestation conditions*. The idea is that these conditions demarcate a set of possibilities – all the possible scenarios where the performance of the action (or manifestation of the disposition) is possible. Manley and Wasserman's account of dispositions quantifies over all such possibilities; I have argued that we normally quantify only over a proper subset of the set of possibilities circumscribed by the minimal action realisation conditions. And this is why we need the idea of a set of *definitional circumstances* which define any particular ability and which will (usually) demarcate some proper subset of the minimal action realisation conditions.

This understanding of abilities enables us to see two ways in which an ability might be general. Distinguishing these two kinds of generality is very important when it comes to the abilities most relevant to free will, as we will see below. The first kind of generality is concerned with the nature of the definitional circumstances. The definitional circumstances demarcate a set of possible circumstances: those which are quantified over in assessing whether this particular ability is possessed. But definitional circumstances *qua* definitional circumstances do not concern the agent's actual surroundings; they are not opportunities. In the following sense, then, the definitional circumstances are hypothetical: if an agent has the ability-to-A-in-C, where C are the definitional circumstances, nothing about the nature of the definitional circumstances themselves entails anything about whether the agent is in circumstances which are of type C. An agent might possess an ability-to-A-in-C and yet not be in circumstances C. In describing the ability's definitional circumstances as hypothetical in the sense just outlined, I do not mean to affirm, as Gilbert Ryle (1967: 41) did, that the ability itself is hypothetical in the sense that ascriptions of abilities do not describe real states which agents are in. An ability is a genuine property which marks a real difference between those agents who have it and those who don't.

However, sometimes in ascribing an ability we do mean to affirm that the agent is in the circumstances relevant to the exercise of that ability; in other words, there is a sense of 'able' which does affirm that the agent has the relevant opportunity. For example, suppose someone asks whether Martha is able to golf. If we answer 'Yes, she's able to golf (very well)', we are naturally taken to be ascribing an ability which says nothing about whether Martha currently has the opportunity to golf. On the other hand, if we answer 'Yes, she's able to golf (her clubs are right here)' we ascribe what Vihvelin would call the wide ability – that is, we ascribe the intrinsic ability to golf and also the opportunity to golf. The first sense – which implies nothing about whether the agent has an opportunity – has often been called the 'general' sense of 'able' or 'ability'; however, I will reserve this name for the second kind of generality to be outlined below. As such I will call these abilities *non-particular abilities*. The kind of ability which does imply the possession of an opportunity I will call *particular abilities*. Thus, if we ascribe to Martha the non-particular ability to golf we say nothing about whether Martha is anywhere near a golf course or has her clubs to hand; if we ascribe to her the particular ability to golf, we do affirm that Martha has the opportunity. In virtue of the role played by the definitional circumstances, the account being proposed has (at least the beginnings of) an account of what it is to possess an opportunity built right into it: to possess an opportunity is to be in circumstances of the same type as the definitional circumstances.

I accept the view, shared by Vihvelin (2013: 7 n.24) and a number of other writers, that the modal claim made by an ascription of an ability may at least be approximated using some 'can' statement: if Martha *is able to* golf, then there is some sense in which she *can* golf. And one natural way of fleshing out this modal claim on the account so far proposed is that it is made relative to or conditional upon the definitional circumstances obtaining. Suppose, for example, that Martha has the ability-to-golf-in-dry-weather (the hyphens, recall, simply emphasise that the circumstances mentioned are part of the characterisation of the ability, rather than being circumstances in which the ability is possessed). Then the modal claim associated with that ability is (roughly): given (or *relative to*, or *conditional upon*) dry weather, Martha can golf. This will be the same for ascriptions of both non-particular and particular abilities. They differ in that particular abilities also affirm that Martha has the opportunity to exercise that ability. On the suggestion made above, which was that to possess an opportunity was to be in circumstances of the same type as the definitional circumstances, this would mean that the particular ability also affirms that Martha is currently in dry weather circumstances. To ascribe a non-particular ability is to say nothing about whether the agent is in circumstances which are of the same type as the ability's definitional circumstances. So if all we say is that Martha has the non-particular ability-to-golf-in-dry-weather then we say nothing about whether Martha is in dry weather circumstances. This explains why, although such an ability ascription will imply that there is *a sense* in which the agent can perform the action, that sense need not mean that the agent can perform the action right now, given the agent's actual circumstances. Non-particular abilities, then, although they are (obviously) possessed in whatever circumstances the agent is in, they are not necessarily applicable to the agent's circumstances because the modal claim they express might be about a set of circumstances which do not obtain. Non-particularity has nothing to do with how much detail is included in the definition of an ability. Consider the following two abilities:

(**A1**) the ability to walk up inclines
(**A2**) the ability to walk up 4% inclines into a head wind of 12 mph

These abilities vary in how much detail the definitional circumstances contain. But both of them could be non-particular abilities.

The second kind of generality, however, *is* highlighted by the difference between **A1** and **A2**. It is very natural to label abilities like **A1** 'general' and abilities like **A2** 'specific'. Here 'general' and 'specific' are being used to characterise the range of circumstances to which the ability applies. Ability **A1** applies to many more circumstances than **A2** and for that reason it is aptly described as general when compared with **A2**. However, while **A2** is properly described as specific when compared to **A1**, it would be general when compared to the following:

(**A3**) the ability to walk up 4% inclines in the rain against a head wind of 12 mph

Used in this way, then, 'general' and 'specific' presuppose some point of comparison with respect to the range of possible circumstances to which the ability applies. I will reserve the term 'general' for this second kind of generality (having called the first kind 'non-particularity'). Generality and specificity then, are a matter of degree: abilities can be more or less general, more or less specific.

What is the relationship between two abilities to A of differing generalities? To answer this question I will turn to the work of Whittle (2010), who put forward an account of abilities which is in some ways similar to that being proposed here. I will begin with a brief overview of her account.

Whittle, unlike most writers on abilities, recognises something akin to what I have called the spectrum of abilities; indeed, the idea of a continuous spectrum of abilities to A is a picture that Whittle explicitly endorses (2010: 8). She constructs her account of abilities by beginning with Manley and Wasserman's (2008) account of dispositions. Manley and Wasserman say that an object possesses the disposition to M when C if and only if that object M-s in a 'suitable proportion' of *all* C-cases, where C represents the stimulus conditions and a C-case (a stimulus condition case) is one fully specified or determinate way that an event of type C might occur (see 2.1.3 for further details). In other words, and as already noted, Manley and Wasserman place no restrictions on the possible circumstances which are salient other than being of type C and sharing the same laws as the actual world.

Whittle thinks that this is a mistake and that to make sense of the variety of abilities we ascribe we need to be able to narrow down the range of possible circumstances quantified over. She therefore proposes an extension to Manley and Wasserman's account which allows C-cases to be restricted in some way. Whittle says that there are two possible kinds of restriction that might be made: we might hold *some but not all* of the factors causally relevant to the exhibiting of the behaviour fixed, or we might hold *everything causally relevant* fixed. And Whittle says that when we restrict some but not all of the C-cases the result will be a *fairly local ability* and when we narrow down to just one C-case then the result is an *all-in local ability* (this distinction applies to both dispositions and abilities). If we don't impose any restrictions then we will have what Whittle calls a *global ability* – this will be the kind of ability that Manley and Wasserman's own account ascribes. For example, the global ability to golf pertains to the entire range of all possible circumstances where an agent tries to golf. A fairly local ability to golf will narrow down to some proper subset of said range; for example, just those circumstances where an agent tries to golf in dry weather. An all-in local ability will narrow down further – indeed, it will narrow down as far as it is possible to go, holding everything fixed that might affect whether the agent succeeds in successfully golfing.

Whittle thus has a threefold classification of abilities: global abilities, fairly local abilities, and all-in local abilities. However, we must be careful not to think that these kinds of abilities differ in the way that, for example, what I've called non-particular and particular abilities differ. Global, fairly local, and all-in local abilities all lie on the same spectrum; in my terminology, they vary according to how much detail is included in each ability's definitional circumstances. Global abilities place no restrictions on the circumstances other than that they are of type C and share the laws of the actual world. Fairly local abilities place some further restrictions on the quantification. All-in local abilities consider just one C-case. In other words, global, fairly local and all-in local abilities vary only according to how general/specific they are and not according to whether they are non-particular or particular abilities – that, at least, is the initial impression Whittle gives of this distinction.

Whereas many writers on abilities recognise something like the non-particular/particular distinction (see 3.3 below) but fail to recognise the general/specific distinction, Whittle tends towards the opposite error. That is, she risks not recognising – or at least not doing justice to – the non-particular/particular distinction. Whittle introduces the distinction between global, fairly local and all-in local abilities by saying that the difference amounts to the range of circumstances to which each kind of ability applies (Whittle 2010: 8). And it's the thought that abilities can apply to different ranges of circumstances that gives rise to the idea that, for each given action type, there is a (continuous) spectrum of abilities – a picture Whittle endorses. This is very similar to the general/specific distinction I've made: as you move from global to all-in local abilities, more and more information is used to characterise the ability, with the result that it applies to fewer and fewer possible circumstances.

But Whittle then confuses things by identifying her global abilities with what Alfred Mele (2003) has called 'general practical abilities' (see below for further details on Mele) – these abilities are characterised by the agent's *lack of any corresponding opportunity* (Whittle 2010: 2). Whittle then contrasts global abilities to local abilities, implying that local abilities require that the agent possess the opportunity. Indeed, Whittle goes on to equate the possession of a local ability with the possession of an opportunity, and to suggest that what I have called the ability's definitional circumstances are equivalent to 'the opportunity to manifest the ability' (Whittle 2010: 8–9). Whittle therefore runs together the two questions that I think it is vital to keep separate: what are the definitional circumstances which partly define the ability? And: is the agent in circumstances which are of the same type as the definitional circumstances? For Whittle it makes no sense to ask whether an agent has the opportunity to exercise a local ability. By contrast, I take the question of whether an agent has the opportunity to exercise a specific ability to be vital. Conflating the two distinctions in the way Whittle does is a major shortcoming; nevertheless, Whittle's discussion of how different abilities which share the same action type but which vary according to their definitional circumstances is still useful.

Perhaps the most important observation Whittle makes is that once we allow the modal force of the disposition or ability to be weaker than a form of restricted necessity, then the possession of two abilities to A of differing generalities can come apart. Whittle gives an example to support this idea (Whittle 2010: 3). Suppose that Sally has a very general ability to sing: she can successfully sing when she tries in a very wide variety of cases. She can sing in such a range of circumstances that she has the most general ability to sing there is – what Whittle calls the global ability to sing. Nevertheless, there is one subclass of ordinary situations in which Sally can't sing: those where her Aunt is present. Sally is intimidated by her Aunt and she freezes up completely when her Aunt is around. So Sally does not have the ability-to-sing-when-her-Aunt-is-present (again, the hyphens indicate that the circumstances mentioned characterise the ability) (Whittle 2010: 3). A similar point holds in the other direction, and Whittle's Sally example can be modified to illustrate this. Perhaps Sally has a terrible fear of social situations, and so could never sing in front of another person. Except, that is, if her supportive and encouraging Aunt is around. With her Aunt around, Sally feels relaxed enough to sing, despite the other people. In this case Sally has the ability-to-sing-with-her-Aunt-present but she does not have the global ability to sing (because this would require singing in many non-Aunt situations).

Whittle provides a number of other examples to help make her case (some of which will be explored below). However, her discussion suffers in the following way. She treats the preceding examples as teaching us something only about the relationship between global abilities – the most general abilities there are – and local abilities and not, for example, as a lesson about the relationship between any two abilities to A which differ in their generality. Moreover, the way Whittle discusses it encourages a further confusion. Summarising her point Whittle writes that 'an agent or object may instantiate the global ability or disposition whilst failing to instantiate *the* local ability' (Whittle 2010: 3; my emphasis). The problem here is that there is not just one local ability – '*the* local ability' – that corresponds to a given global ability. Global abilities lie at one end of the spectrum: they are not different in kind to local abilities; rather, they differ simply because they define the ability using less information (as little as is possible, for the given action type). In my terminology, global abilities are the most general abilities. I will call these *maximally general abilities*. There is only one maximally general ability for each given action type; e.g. the maximally general ability to sing. But there are many, many local abilities for each given action type – hence the spectrum of abilities. So we have, for example, the ability to sing in front of strangers, the ability to sing in front of one's family, the ability to sing in front of one's Aunt, and so on. Whittle agrees that there is such a spectrum of abilities; what she appears to overlook is that this means there is no one local ability to A which corresponds to the global ability to A.

Why think that the kind of conclusions Whittle draws might apply, not just to some given global ability to A and any local ability to A, but also to any two abilities to A which differ with respect to their generality? Well, there are two reasons that the possession of the global ability to A and the possession of some local ability to A might come apart. First, the global ability to A quantifies over a different set of possibilities to any local ability to A. Second, the modal force is not a form of restricted necessity. But these things also hold for any two local abilities to A. In my terminology, the point holds for any two abilities to A regardless of their generality. To illustrate with Whittle's Sally example. Whittle singles out Sally's ability to sing in front of her Aunt as *the* pertinent local ability. But of course, the ability to sing in front of one's Uncle is another local ability that lies on this same spectrum. Sally's possession of the global ability to sing need not imply her possession of this latter local ability either. And again, we could also ask about the ability to sing in front of one's Aunt and Uncle. Perhaps Sally can sing in front of her Aunt, and she can sing in front of her Uncle, but if both her Aunt and Uncle are present it's just too much for her. Only then does she fail in her singing. So Sally might possess the first two local abilities just mentioned – the ability to sing in front of her Aunt and also the ability to sing in front of her Uncle – but not possess the even more local (i.e. even more specific) ability to sing in front of her Aunt and Uncle. The correct lesson to draw from Whittle's observation is therefore as follows: for any given ability to A there is not a certain more specific ability to A which *must* be possessed.

Is it the case, nevertheless, that if an agent possesses a non-maximally specific ability to A, then they must also possess some or other more specific ability to A? It might seem like it, for an agent will only possess the ability to A-in-C if there are some C-cases in which the agent succeeds in A-ing. And it might be thought that we can simply take some feature possessed by those possible C-cases in which the agent succeeds in A-ing but not by those in which the agent fails, add that feature to the definitional circumstances, and we will end up with a more specific ability to A that the agent will possess: the ability-to-A-in-C-and-D, where D is the feature belonging only to those C-cases in which the agent succeeds. But this assumes that all the cases where the agent succeeds share some feature that is not had by any of those cases where the agent fails, and that need not be so. Still, it might be suggested that we can arrive at a more specific ability by adding some feature possessed by any (but not all) of those cases where the agent succeeds. So we would have the ability-to-A-in-C-and-E where E is a feature had by one or more cases in which the agent succeeds and none of those cases where the agent fails. In many cases there will be such a more specific ability. But even this is not guaranteed; if, for example, the universe is indeterministic, then there need be no such feature.

What about the converse: if someone possesses an ability to A does it follow that there is some or other more general ability to A which they must possess? In some cases this does appear to be so; consider the case where Sally has the ability to sing in front of her Aunt. Then it might be suggested that Sally must also have the ability to sing in front of someone who is supportive in ways X, Y, Z (where X, Y and Z are the features of Sally's Aunt which make her supportive), and this ability looks to be more general than the ability to sing in front of her Aunt. But it's not clear that this is always the case. Take, for example, the ability just mentioned: the ability to sing in front of someone supportive in ways X, Y, Z. If X, Y, Z are purely qualitative features of a person, there may not be any ability to sing more general than that which the possession of said ability would imply having.

This conclusion – that possession of an ability to A does not imply the possession of any more general ability to A – holds as long we make the simplifying assumption that there is no variation in the modal force of all abilities-to-A. However, I have argued that abilities are in part individuated by the modal force. Thus, for a given action type, A, there is no modal force value that all abilities to A must have. The modal force, just like the definitional circumstances, can vary. For example, two people might have the ability-to-sing-in-front-of-a-large-audience but to different degrees. Sally, for example, might be able to sing (when she tries) in, say, 95% of all the cases involving large audiences, while Samuel might be able to sing only in 70% of such cases. Sally and Samuel thus have different abilities – and the abilities differ only in their 'strength', only in their modal force.

Once we allow the modal force of the abilities to A to vary it will be the case that possessing some ability to A implies possessing some or other more general ability to A. This is because, in order to accommodate the wider range of circumstances which are associated with a more general ability, we can always reduce the modal force. For example, if Sally has the ability to sing in front of a small audience with a modal force of 95%, then she will also have the ability to sing in front of an audience (probably with some lower modal force). The latter ability includes cases where she's in front of a huge audience. But even if Sally always fails in such cases, still, the modal force can be lowered accordingly.

There is one final consideration concerning the specification of abilities that it is important to highlight. The circumstances which define an ability are restricted in certain ways by what I have called the *minimal action realisation conditions*. Recall that these circumstances demarcate the widest possible set of scenarios for which an action of a given type is a possibility. Outside of this set, the action is not possible (no action in circumstances outside this set would count as being an action of the relevant type). This means that an ability's definitional circumstances cannot include circumstances which fall outside the minimal action realisation circumstances. Suppose we accept, for sake of illustration, that walking requires a hard surface. Then the action realisation circumstances for walking demarcate a set of possible scenarios where there is a sufficiently hard surface. Given that, if we were to specify an ability to walk which includes, as part of its definitional circumstances, the stipulation that there is no hard surface, then we would have an ability that it would be impossible to possess.

3.3. The kinds of ability relevant to free will

We are now in a position to answer some aspects the following question: which kind(s) of ability are relevant to free will? In the previous section I drew two distinctions: a non-particular/particular distinction and a general/specific distinction. This gives us (at least) two questions. First: does free will require non-particular abilities or particular abilities? Second: does free will require general or specific abilities? Both of these questions presuppose that free will always requires the same kind of abilities, and in what follows we will see that there are reasons to accept this. Indeed, this view is widely shared among those who agree that free will requires the ability to do otherwise. It is common among such theorists to articulate a distinction between those abilities not relevant to free will and those that are; typically this involves contrasting some notion of a general ability with some notion of non-general ability. But although the same kind of examples are appealed to in support of these distinctions, the distinctions drawn are often incompatible. I want to suggest that this is because the two different ideas lying behind the term 'general' – in my terminology, non-particularity and generality – are often conflated. I will briefly illustrate this point.

Mele characterises general abilities as those we attribute 'even though we know [the agent] has no opportunity to [exercise the ability] at the time of attribution' (Mele 2003: 447). He contrasts these with specific abilities which are those an agent has at a certain time to perform an action *then or on some specified occasion*. Mele's specific abilities are specific because the *exercise* of the ability is tied to a particular time – either the time at which the ability is being ascribed or shortly afterwards. If the potential exercise of the ability is tied to a time in this way – if it is possible that the ability be exercised at that time – then it appears safe to conclude that the agent has the opportunity to exercise that ability (especially given Mele's characterisation of general abilities as those possessed without opportunities). Mele's specific abilities are therefore very similar to what I have called particular abilities: they are possessed when the agent possesses a non-particular ability to A-in-C and also the opportunity to A-in-C. His classification contains no recognition that both non-particular and particular abilities (as I call them) are subject to a further classification along the lines of the general/specific distinction. This makes Mele's classification very similar to Vihvelin's (as discussed in the previous chapter). The main difference between them is that Vihvelin is explicit in saying that what she calls narrow abilities (Mele's general abilities) are intrinsic properties whereas wide abilities (Mele's specific) are extrinsic – Mele does not explicitly mention the intrinsic/extrinsic distinction.

Mele's specific abilities (and Vihvelin's wide abilities) stand in sharp contrast to Ferenc Huoranszki's (2011) specific abilities. Huoranszki thinks that almost all of an agent's abilities, and certainly all those relevant to free will, are extrinsic. But they are not extrinsic because they require the possession of an opportunity, as with Mele's specific abilities and Vihvelin's wide abilities. No, according to Huoranszki an agent might possess the abilities required for free will without possessing any associated opportunity (Huoranszki 2011: 32, 86). Rather, Huoranszki takes abilities to be extrinsic because he thinks actions themselves have to be extrinsically identified (Huoranszki 2011: 62ff). For example, to murder someone is (roughly) to kill someone unlawfully; thus, Huoranszki suggests, no one can have the ability to commit a murder in the absence of whatever societal structures give rise to the relevant laws. According to Huoranszki, then, the ability to murder is always extrinsic. Huoranszki thinks this point generalises; e.g. Abby hands Becky a small, round metal disc; to adequately characterise what Abby has done we need to make reference to what money is, how it is used, whether Abby owes Becky any money (Abby is repaying a debt), whether Becky has something Abby wants and is willing to exchange it (Abby is buying something), whether Abby wants Becky to behave in a non-lawful manner (Abby is bribing Becky), and so on (this example comes from Shaun Gallagher and Dan Zahavi (2008: 157)). To adequately characterise what Abby is able to do, then, we need to make reference to many things which are extrinsic to Abby. A detailed consideration of Huoranszki's argument would take us too far afield. The crucial point is that in contrast to the *possession* of the ability, which Huoranszki always takes to be an extrinsic matter, he thinks that whether or not an agent *exercises* that ability always depends on conditions intrinsic to the agent (Huoranszki 2011: 62). His position is therefore the reverse of the more common stance found in Mele and Vihvelin where the agent's possession of the ability depends on intrinsic conditions and the agent's exercise of the ability depends (at least in part) on extrinsic conditions.

Both the Mele/Vihvelin classification and Huoranszki's classification are incompatible with a distinction drawn by Antony Honoré (1964) between a general and a particular sense of 'can'. Honoré's general sense of 'can' is similar to what Mele calls general abilities (my non-particular abilities). To say that an agent can A in the general sense, is to say nothing one way or the other about whether the agent is going to A or even going to try to A (Honoré 1964: 465). And nothing is said one way or the other about whether the agent has an opportunity to A. Rather, what's being said is that *on the assumption that the agent tries* the agent would (probably) succeed. The parallels here are obvious. What's interesting – and what makes Honoré's account incompatible with all of the classifications considered so far – is that his understanding of the particular sense of 'can' is non-modal. The particular sense of 'can' Honoré identifies is tied to whether the agent does or does not perform the action. He gives good evidence that there is such a sense of 'can'. For example, the sentence 'I can see you in the undergrowth' is roughly equivalent to 'I *am seeing you* in the undergrowth'. It makes no sense to ask someone who has just made such an utterance, 'Yes, but did you see me?'. This sense of 'can' is tied to what the agent is or is not doing. The ability to do otherwise cannot, therefore, involve Honoré's particular sense of 'can'. Whittle's classification, already encountered and to be discussed further below, is different from all of the above in that she recognises what I have called the general/specific distinction (although as mentioned she conflates it with the non-particular/particular distinction).

All of the authors just mentioned, however, claim that the distinctions they draw are useful – some say crucial – to understanding the abilities involved in free will. All draw the distinctions they do in order to articulate something about the sense of 'can' or 'able' that is relevant to free will. In some cases, and despite the incompatibilities, an author explicitly identifies the distinction they draw with one or more of those drawn by other authors (e.g. Whittle (2010: 2–3) equates her distinction with that drawn by Mele). One of the contentions of this chapter is that the accounts outlined above – and the discussions of free will which then ensue – differ so much because there are in fact two distinctions at work, namely, those two distinctions already outlined: non-particular vs particular abilities on the one hand, general vs specific abilities on the other. In what follows I will endeavour to keep these questions separate; I will address the non-particular/particular distinction first, followed by the general/specific distinction.

3.3.1. Does free will require non-particular or particular abilities?

Consider the first question posed above: does free will require non-particular or particular abilities (or perhaps either will do)? Put another way: does free will require the possession of a non-particular ability *and also* the opportunity to exercise that ability, or does the possession of a non-particular ability suffice? There is very wide agreement that the sense of 'able' relevant to free will expresses the idea that, given an agent who has performed an action, that agent had not only the non-particular ability to do otherwise but also the opportunity to do so. Mele dismisses the relevance of what he calls general abilities in the first few paragraphs of his discussion (Mele 2003: 447). The kind of ability relevant to free will, according to Mele, is a species of specific ability which requires possessing the opportunity. Vihvelin thinks the same ((Vihvelin 2013: 7 n.26); see also 2.1.4). This also appears to be true of Whittle. Although her spectrum of abilities bears some similarity to the general/specific distinction I have drawn, as already argued, Whittle conflates this distinction with the non-particular/particular distinction. And, at least at some points in her discussion, Whittle appears to say that local abilities are the ones relevant to free will because they involve the possession of an opportunity. (I will show in the following section that Whittle's main arguments for the need for local abilities seem to revolve around features of their specificity, rather than their particularity). Of those writers already mentioned, Huoranszki is the only clear dissenter to the view that free will requires the possession of an opportunity (Huoranszki 2011: 86).

Many writers who do not explicitly endorse any method for classifying abilities also agree that opportunities – or at least something very like them – are required. A. J. Ayer (1977: 318) and P. H. Nowell-Smith (1960: 101) both say that the agent is able to do otherwise in the relevant sense only if some conditions extrinsic to the agent are met. The extrinsic conditions outlined are naturally thought of as amounting to, or themselves requiring, the possession of an opportunity – Ayer, for example, requires that the agent's environment be such that the agent's choice would cause the agent's action and also that there be no other people compelling the agent to make any particular choice. Similarly, Susan Wolf's account of the relevant kind of ability requires that there be nothing extrinsic which would 'interfere with or prevent the exercise' of the agent's capacities, skills and talents (Wolf 1990: 101).

Despite the wide consensus that opportunities are required, few if any of the above authors provide any explicit argument to this effect. For the most part intuitive considerations about the need for an ability to be exercisable – what good would it be to have the ability, if it weren't exercisable? – and the benefits of not being compelled or constrained do most of the work. These are of course important considerations. But I want to suggest that the account of abilities proposed here gives us a further reason for taking opportunities to be important, namely, that an ability apply to the agent's circumstances in the first place. Whether an ability applies to the agent's current circumstances is closely connected to whether it is exercisable; but the two issues do come apart: an ability which applies to the agent's current circumstances might yet not be exercisable (an ability which does not apply to the agent's current circumstances will of course never be exercisable). The current proposal gives us a way of thinking clearly about these issues. Abilities are partly defined by a set of definitional circumstances – those circumstances to which the ability applies. And opportunities are tied to individual abilities: an agent has an opportunity to A-in-C if that agent has the ability to A-in-C and is in circumstances of type C. To see the importance of opportunities according to this account, suppose that I have the following list of non-particular abilities (the circumstances mentioned characterise the ability, but I omit the hyphens for readability):

(**A4**) The ability to swim in calm waters
(**A5**) The ability to juggle with 5 balls when standing on a flat surface
(**A6**) The ability to jump 75cm off the ground
(**A7**) The ability to sing when standing at rest

Suppose, in addition, that I am in the middle of a desert; there is no body of water for miles, nor any juggling balls. Intuitively, only abilities **A6** and **A7** say anything about what I can do in my current situation. After all, in the above scenario, it is up to me whether I jump 75cm, and up to me whether I sing, but it's not up to me whether I swim or juggle. Why is that? It's not that abilities **A4** and **A5** *don't* express a claim about what I can do; they do. Each says that something is possible for me. And the possibility claims they entail are true of me even while I'm stood in the desert. That is, in virtue of possessing **A4** – the ability to swim in calm waters – it is true to say of me, even while I'm stood in the desert, both that I can swim in calm waters and that it is possible for me to swim in calm waters. But this means that there is *a sense* in which I can swim. Why is that not relevant to whether swimming is up to me? The reason is that these possibility claims express nothing about what I can do *given my current situation*. The kind of possibility involved here, as hinted at above, is a kind of *relative* or *conditional* possibility where what is said to be possible – my swimming – is said to be possible *relative to* or *conditional upon* circumstances which do not obtain. The reason **A4** and **A5** are irrelevant, then, is not (just) that the associated possibility claims are abstractions – i.e. that they affirm the possibility of something only relative to some subset of the set of facts which completely describe the universe. The main problem is that the possibility claims entailed by abilities **A4** and **A5** are made relative to a set of propositions not all of which are true: the set includes propositions describing non-actual states of affairs. To say that the sense of 'able' relevant to free will must express the idea that the agent possesses an opportunity is thus to say that the set of propositions that the action is said to be possible relative to must all be true. Being clear about the role that definitional circumstances play in defining abilities, and therefore the role definitional circumstances play in characterising opportunities, helps us to be clear about this point.

None of this is to say that the presence of an opportunity guarantees that an ability is exercisable. This is for two reasons. First, there are many non-particular abilities to A, and so there are many particular abilities to A. This means that an agent might have many particular abilities to A all of which have definitional circumstances of the same type as the agent's actual circumstances. On the current account, then, demanding that an ability be a particular ability is only the first step in determining whether an ability is relevant to free will. Second, and related to the point just made, abilities can be masked. This is possible precisely because most abilities are characterised by some level of generality which opens up the possibility of interfering factors that might inhibit the exercise of the ability. Both of these issues are concerned with the general/specific distinction and will be dealt with below.

It might be objected that the account proposed is not quite right because it cannot make sense of scenarios where the agent has the ability to *try* to do something despite lacking the ability to do it. And inasmuch as I have argued that the sense of 'able' relevant to free will must express the idea that the agent is able to try (or choose or begin) to perform the action in question, I will need to deal with cases where the agent is able to try but not able to succeed. This is especially true given that it is plausible to think an agent's being able to try (but not succeed) is often enough to give the agent a significant level of free will (a claim for which I will argue below).

The thought could be developed as follows. Suppose that I'm stood in the middle of a desert and have both the non-particular ability-to-swim-in-calm-waters and the non-particular ability-to-*try-to*-swim-in-calm-waters. Although I cannot (successfully) swim in calm waters when I'm in the desert, I can try to do so; I might not believe, for example, that I'm in the middle of a desert. Moreover, this seems to reveal a potentially significant option that is available to me: I can try to swim. If I can try to swim, then it might be suggested that I can freely refrain from swimming. At the very least, if I'm able to try to swim and able to refrain from trying to swim, then it seems that I have free will with respect to my trying to swim.

But, the objector might continue, according to my proposal this cannot be explained. Both the ability-to-swim-in-calm-waters and the ability-to-try-to-swim-in-calm-waters, on the view I've proposed, apply only to situations where there is a body of calm water available. That is, according to the objector, the definitional circumstances for each ability are *being in calm waters*. But given that I'm not in calm waters, then I can possess only the non-particular ability to try to swim in calm waters. I can't possess the particular ability to swim in calm waters because that would require having the opportunity to exercise that ability, which would mean actually being in calm waters. According to the objector, then, my account cannot explain my ability to try to swim when in a desert.

This kind of objection raises a number of important (and difficult) issues. First of all, the claim that I can try to swim in a desert could be questioned. To the extent that agency is embodied, the movements that make up a successful swimming stroke, front crawl, say, will be a function of my trying to front crawl *and my actually being in water*. The movements I would end up making if I were to try to front crawl while lying on the sand in a desert would be quite different to those which I make when successfully swimming. If this is not obvious, consider what it is like to walk on a bouncy castle for the first time. If you were to attempt to 'walk normally' (as if walking on a pavement, say) in a bouncy castle, the movements you would end up making would be very different to those produced when you are in fact on a pavement.

However, I will put this issue to one side. There are scenarios where an agent might try to swim in a desert – the agent might believe there is water there, or not believe they are in a desert – so it would be good if my account had some way to explain what goes on in these cases. What I want to suggest is that trying to swim is a different action type to swimming. It therefore has different action realisation conditions associated with it. In particular, whereas the minimal action realisation conditions for swimming require being in a suitable liquid, the minimal action realisation conditions for trying to swim do not. That is why we can ascribe an ability like the ability-to-try-to-swim-in-a-desert. Indeed, we could even ascribe the-ability-to-try-to-swim-in-calm-waters-in-a-desert, if *swimming in calm waters* is understood as being the full characterisation of the action type. In other words, the qualification 'in calm waters' attaches to the action type and not to the definitional circumstances. We don't usually make such fine distinctions, distinguishing between, say, the actions of *swimming in calm waters* and *swimming in choppy waters*. But there are contexts where such distinctions are made. A seasoned open water swimmer might develop a technique for swimming outside which differs to that used when swimming in a pool. Such a swimmer might think of these as two slightly different, although closely related, activities. And so such a swimmer could try *to swim in choppy waters*, where all of that description pertains to the content of the swimmer's intention (and not to the swimmer's environment).

My account, then, can make sense of abilities which appear to have contradictory definitional circumstances – being in calm waters and being in a desert – by construing some part of the characterisation as being an addition to the action type. Of course, this does not work for abilities to *perform* those actions. This is because the action type puts constraints on the definitional circumstances. But this gives just the right results. I do not – and cannot – have the particular ability to swim-in-calm-waters-in-a-desert, but I do have the particular ability-to-try-to-swim-in-calm-waters-in-a-desert (provided that is understood in the way described above).

What about the options that this ability to try to swim gives me – are they significant? If I can try to swim in the desert, does that mean I can freely refrain from swimming? No. I cannot freely refrain from doing something which it is not possible for me to do. However, I can freely refrain from trying to swim. This is something I can do freely, on the account being proposed, because I have both the ability to try to swim and the ability to refrain from trying to swim. Thus, it is up to me whether or not I try to swim. And so when stood in a desert I have free will with respect to performing an action of type *trying to swim* but not an action of type *swimming*. Of course, I might not know that I cannot swim in the desert. I might have the mistaken belief that it is possible to swim in air, say, or in the sand. But all this illustrates is that I might have less free will than I think I have.

3.3.2. Does free will require general or specific abilities?

What should we say about the second distinction, i.e., the distinction between general and specific abilities? Which ones are relevant to free will? As posed the question is ill-formed because abilities are not general or specific *simpliciter*, but only when compared to another ability. For any given action type, A, there is a spectrum of abilities to A which vary in how general or specific they are according to the level of detail in the definitional circumstances. The most obvious question to ask instead is whether there is some fixed level of specificity which is always required for free will. I will argue that the answer to this question is yes, and that the level of specificity we're looking for is that associated with the most specific abilities, which I will call *maximally specific abilities*.

In answering this question there is not much help available in the literature. As already noted, many of those philosophers who do discuss something they call the general/specific distinction are referring to what I have labelled the non-particular/particular distinction (Mele was an example of this). The one writer who does discuss these questions in some depth, however, is Whittle. I will therefore begin with her discussion.

Whittle makes two claims with respect to which abilities are relevant to an agent's possessing free will. The first claim is that free will does not involve global abilities – in my terminology, it does not involve maximally general abilities. To support this idea Whittle provides two examples which are meant to show that fairly local abilities are more relevant to an agent's free will than global abilities – indeed, she claims that her examples show that global abilities are not just less relevant but irrelevant to an agent's free will. Whittle's second claim is that nothing about the examples she presents lead us to think that free will might require all-in local abilities. In my terminology, Whittle's position is that free will requires abilities that are neither maximally general nor maximally specific: the relevant abilities lie somewhere in the middle of the spectrum. The first of Whittle's examples is as follows:

> (**Bound Ben**) Ben, an excellent swimmer, has been forcibly bound to a chair. He watches helplessly as a child drowns in a lake (Whittle 2010: 10).

Whittle says that in this example Ben has the global ability to swim – the maximally general ability to swim, using my terminology. Call this ability **A8**. Intuitively, despite possessing this ability, Ben is not morally responsible for the child's death. This suggests that 'global abilities do not connect up with our intuitive judgements regarding moral responsibility' (Whittle 2010: 12). Why does Whittle appeal to moral responsibility in order to establish the irrelevance of global abilities for free will? She doesn't make her reasoning explicit, but presumably it is something along the following lines. In the example we are assuming that Ben is an ordinary, mature human adult, such that other things being equal he is morally responsible for his actions. And Whittle assumes – as I do – that being morally responsible requires that one have and exercise free will. On the current assumptions, this means that to be responsible, Ben must perform some action and be such that he was able to have done otherwise. If, therefore, we have an example where Ben possesses one kind of ability but lacks another, and if all other things are equal, then a good explanation for Ben's lack of moral responsibility is that he did not have free will with respect to saving the child. From this it follows that the kind of abilities Ben *does* possess – global abilities – do not suffice for free will. This, I suggest, is why Whittle comes to the following conclusion:

> Given that Ben is bound to a chair, what we should be considering when judging whether he was morally responsible for the child's drowning is whether he instantiates the fairly local ability-to-swim-when-bound-to-a-chair (Whittle 2010: 10).

Whittle's second example illustrates the reverse point:

> (**Obedient Olive**) Olive has been conditioned in a concentration camp. She receives orders specifying what she should do every time she comes to make a decision, and she is unable to do other than what she is instructed to do. Except, that is, when she gets a rare instruction from Derek. Then it is up to her (and Olive is aware of this) whether or not she obeys the orders. One day Derek tells her to smash some windows and Olive does so (Whittle 2010: 12).

According to Whittle, in this example Olive lacks the global (maximally general) ability to do other than she is ordered to do. That's because in most cases when she faces a number of options there is only one thing she can do, namely, that which she's instructed to do. But, Whittle says, it is plausible to say that lacking this global ability does nothing to pardon her for smashing the windows – we should hold her morally responsible for that, despite her not being able, in general, to disobey her orders. This is because in virtue of having the fairly local ability-to-do-otherwise-given-Derek's-orders, Olive appears to have all the control required to bear responsibility for smashing the windows. She was able to follow Derek's orders and was able to refrain from following Derek's orders; it was therefore up to Olive whether she followed Derek's orders – she had free will with respect to smashing the windows (Whittle 2010: 12). Whittle therefore concludes that because Olive lacked the global ability to do otherwise, global abilities are irrelevant to free will.

Whittle's examples are useful, and I accept as legitimate the method of inquiry whereby we consider scenarios where it's intuitive to think of someone as morally responsible (or not) and then ask what best explains this. Ultimately, however, Whittle does not apply this method correctly and so draws erroneous conclusions from her examples.

Remember, first of all, that for each action type there is a spectrum of abilities, global at one end and all-in local at the other. In my terminology, there are maximally general abilities at one end and maximally specific abilities at the other, with a huge number of abilities in between (Whittle's fairly local abilities). I will call these abilities *non-maximal*: they are neither maximally general nor maximally specific. The vast majority of abilities to A will be non-maximal. After all, for each action type A, there is only one maximally general ability to A – it quantifies over the set of possible scenarios demarcated by the minimal action realisation conditions for A. As soon as some further restriction is placed on the set of possibilities to be quantified over we have moved away from a maximally general ability – this is the case no matter how small that further restriction is. And unless we move right along the spectrum of abilities such that we end up specifying everything that is causally relevant to the performance of the action, we still have a non-maximal ability. Most abilities are therefore non-maximal. Why is this important? Well, it means that Whittle's claim that fairly local (i.e. non-maximal) abilities are more relevant than global abilities is not very substantive. For each action type, A, Whittle's claim rules out precisely one ability as being the one required by free will.

Whittle goes on to claim that 'nothing said here' – i.e. nothing about her examples – provides any reason to think that it will be all-in local abilities that are relevant to free will (Whittle 2010: 18). So Whittle takes her examples to support the idea that non-maximal abilities trump both maximally general abilities and maximally specific abilities as far as being relevant to an agent's free will is concerned. Whittle gives no argument for the second part of her claim; she simply makes the bald assertion that the examples presented don't lead us to maximally specific abilities (i.e. her all-in local abilities). I will challenge the second part of her claim below, but it is worth nothing that even if we were to accept that maximally specific abilities are not required, Whittle's discussion would still be incomplete. She says nothing, for example, about *which* of the fairly local abilities are relevant. Remember, although there is only one maximally general ability for each given action type, A, there are a vast number of non-maximal abilities to A. And despite saying that she envisages a 'continuous spectrum' of abilities to A, from global to all-in local, she routinely writes as if there is only ever one fairly local ability for each action type; she states, for example, that 'an object can have both *the* local disposition ... and *the* global disposition, or it can have either one without the other' (Whittle 2010: 7). This is a mistake. There is no privileged non-maximal ability (local ability, in Whittle's terminology) corresponding to each maximally general ability. So Whittle's discussion is incomplete, at best.

I return now to the second of Whittle's claims, namely, the idea that there is nothing in her examples which leads us to conclude that it is maximally specific (all-in local) abilities that are the abilities relevant to free will. Whittle, I would like to suggest, fails to draw the correct conclusions from her examples. In a sense, the examples do support the claim that fairly local abilities are more relevant to free will than global abilities. But that's because they support the more general claim that, given any two abilities to A, the more specific will be more relevant to free will. We can extend Whittle's own examples to illustrate this point. Consider again the **Bound Ben** example. Whittle says that it is the fairly local ability to swim-when-bound-to-a-chair that we should think of as relevant to Ben's free will. If he had that ability – if he was 'fairly locally' able to swim when bound to a chair – then we could conclude that it was up to him whether he saved the child, and so we could conclude that he is morally responsible. But is that fairly local ability really the last word on whether Ben could have saved the child? To see why it is not, consider the following example:

> (**Bound Benjamin**) Benjamin, an excellent swimmer, has *had both legs and both arms* forcibly bound to a chair. He watches as a child drowns in a lake.

Suppose we ask whether Benjamin is responsible for failing to save the child. Should we identify, as the ability relevant to determining whether Benjamin acts with free will here, the ability that Whittle appealed to in the case of Ben? That is, should we ask whether Benjamin has the ability-to-swim-when-bound-to-a-chair (call this ability **A9**)? Or should we ask about an ability which includes as part of its characterisation the information mentioned in this case, i.e., the ability-to-swim-with-both-legs-and-both-arms-bound-to-a-chair (call this ability **A10**)? This is not an idle question. The two abilities differ in how general they are and Benjamin's possession of **A9** and **A10** might come apart in the way that Ben's possession of **A8** (the global ability to swim) and **A9** came apart. That is, just as Ben possessed **A8** but not **A9**, Benjamin might possess **A9** but not possess **A10**. He might possess **A9** because he can swim in a enough of the **A9** cases – the 'being bound to a chair' cases – to count as having that ability; after all, the 'being bound to a chair' cases include those cases where an agent is bound, say, just by one arm, or just by one leg. Benjamin might well be such a strong swimmer that if he's got two or three limbs free he can swim even when he has to pull a chair along behind him. Still, even if he possesses **A9** he might lack **A10**: despite his strength, swimming when all four of his limbs are bound to the chair is just too much.

Benjamin's lacking **A10** therefore seems to be more relevant to his free will and moral responsibility than his possessing **A9**. But if that is so, then contrary to what Whittle claims, the same is true of Ben. To see this, note that the two cases are compatible: **Bound Benjamin** might just be a filling out of the **Bound Ben** case. Ben, in other words, might be bound in exactly the way that Benjamin is bound; Ben might be Benjamin. Crucially, it appears that this line of reasoning could be applied again and again, for increasingly specific abilities. This will ultimately lead us to the maximally specific abilities as those which are relevant to free will. Suppose, for example, that Benjamin is indeed bound to the chair by all four of his limbs – *but only with a single strand of hair*. His having all four limbs bound to a chair will now not be much impediment to his swimming. But note that Benjamin will still lack the ability to swim with all four limbs bound to a chair (ability **A10**). The ability he possesses – and the ability that seems to render **A10** irrelevant – is the ability to swim with four limbs bound to a chair with a single strand of hair (call this **A11**).

Note too that it is Benjamin's *possession or lack* of this more specific ability (**A11**) that matters. It is not as if, were Benjamin to lack **A11** – suppose he has a pathological aversion to hair, similar to Lehrer's pathology of red candy (discussed in 1.2.3), such that he cannot even try to move when bound with hair, even just a single strand thereof – the slightly more general ability to swim when bound to a chair which he does possess (i.e. **A9**) would suddenly become relevant to his free will. No, if Benjamin lacks **A11** – and assuming, of course, that he is in circumstances of the same type as **A11**'s definitional circumstances – then he is unable, in the sense relevant to free will, to swim.

If this is correct, then it is not quite right to say that for any two abilities to A, the more specific ability is *more relevant* to free will than the more general. Strictly speaking, the only abilities relevant to free will are maximally specific abilities. What is correct, however, is that given any two abilities to A, the more specific ability will always be a better guide to whether the agent has free will with respect to A than the more general ability. It is also worth noting that although the abilities relevant to free will are always maximally specific, this doesn't mean that all of an agent's maximally specific abilities are relevant to their free will on every occasion. We need to combine the conclusion just drawn with the conclusion drawn in the previous section, namely, that the only abilities relevant to an agent's free will at any given time are those the agent has the opportunity to exercise. Benjamin, for example, will have lots of a maximally specific abilities to swim, it's just that none of them are also particular abilities; none of them are abilities he has the opportunity to exercise. What is needed for someone to have free will with respect to the performance of an action of type A, then, is that the person have a particular, maximally specific ability to A and a particular, maximally specific ability to do something else. The latter may be nothing more than an ability to refrain from A-ing, but it may be, say, the ability to make a different choice or to perform some different bodily action.

3.4. Abilities to do otherwise and Vihvelin's treatment of Frankfurt-style cases

Vihvelin does not recognise the existence of a spectrum of abilities properties for each given action type. As a result she does not address the kind of questions raised in sections 3.2 and 3.3. For any an action type, A, Vihvelin recognises only three kinds of ability: the skill to A, the narrow ability to A and the wide ability to A. As previously discussed, the narrow ability (and therefore the wide ability) is tied to some set of relevant test-cases which are determined by applying the 'getting specific strategy' (see 2.1.3 and 2.3.2 for details). Putting skills to one side, and adopting my terminology, we could say that Vihvelin recognises the following two abilities: the non-particular ability-to-A-in-the-relevant-test-cases and the particular ability-to-A-in-the-relevant-test-cases.

Why does this mean that Vihvelin's account of the ability to do otherwise is inadequate? Well, the only way Vihvelin's account could prove to be satisfactory is if, out of the spectrum of abilities, it were always the ability-to-A-in-the-relevant-test-cases that were relevant to free will. If this were the case then although Vihvelin never addresses the questions raised in 3.2 and 3.3 it would not matter. She would have 'lucked out' as it were: she failed to address some crucial questions, but the account she presented turned out to include the correct answers to those questions anyway. But as the results of section 3.3 show, this is not the case. Free will requires particular, maximally specific abilities. As Vihvelin envisages the set of relevant test-cases, the associated abilities will not be maximally specific. So Vihvelin's account of the ability to do otherwise picks out the wrong kind of ability.

The problem could be put like this. Once we recognise the spectrum of abilities, we see that there are many different ways that an agent may be able to do otherwise. This is not just because it's possible that an agent might be 'non-particularly able' to do otherwise and 'particularly able' to do otherwise. That distinction is important, but there are also many ways of possessing the non-particular ability to do otherwise and many ways of possessing the particular ability to do otherwise. And not every way of being able to do otherwise is relevant to an agent's free will. Vihvelin's account of the ability to do otherwise is an account of just one way that the agent may be able to do otherwise, but it's not a way of being able to do otherwise that matters for free will. I will show how this undermines Vihvelin's treatment of the Frankfurt-style cases.

Recall that Frankfurt-style cases are scenarios which purport to show that an agent can be morally responsible – and so can possess any free will required for moral responsibility – despite not being able to do otherwise.

Typically they present us with an agent, Jones, who performs an action 'on his own'. Lurking in the background is a meddling neuroscientist, Black, who wants Jones to pursue a certain course of action. We are to imagine that, on some particular occasion, it turns out that Jones does 'on his own' exactly what Black wants him to do. Crucially though, Black can detect when Jones is going to deviate from his plans and he can interfere with Jones – usually via brain manipulation – to cause Jones to behave as he wants. Black can therefore ensure that Jones acts in a certain way. But because Black's intervention was not needed, it's claimed that Jones is responsible for his action and therefore must have been free. If this is right, then free will does not require the ability to do otherwise.

Vihvelin, of course, thinks that the ability to do otherwise *is* required for free will so it behoves her to say something about such cases. She begins by dividing the Frankfurt-style cases into two kinds according to how the potential intervention would be achieved, if it were needed. Black, Vihvelin says, is either a 'Bodyguard' or a 'Pre-emptor'. The difference is as follows. Black needs to be able to tell whether his intervention is necessary. To do this, Frankfurt-style cases can either have Black's intervention *conditional* upon Jones's *beginning to deviate* from Black's plans, or they can have Black's intervention triggered by some *perfectly reliable 'prior sign'*, a sign which occurs before Jones acts in any way but which indicates what Jones will do. In the first kind of case Black is a Bodyguard: he curtails the range of things that Jones might successfully do by standing ready to forcibly stop Jones doing anything different the moment he sees Jones step out of line. In the second kind of case Black is a Pre-emptor: Black detects what Jones would do – his own intervention aside – before Jones begins to do it; if he detects non-compliance, he manipulates Jones's brain to ensure that he does what Black wants. Vihvelin treats these kinds of case differently; I will address each in turn.

Consider the case when Black is a Bodyguard. This is the more straightforward case because there is no mystery as to how such intervention is possible: Black simply needs to be able to detect the beginnings of Jones's action. Black can detect, we are to suppose, even the very first beginnings of a decision. So if Jones were to begin to decide in a way that Black does not want, Black would intervene right away and 'fix things' such that Jones makes the 'right' decision. There might be a 'beginning of a decision' that is not to Black's liking, but there will never be a completed decision which is contrary to Black's designs. According to Vihvelin, this means that Jones is only able to perform one type of action – and this is the case whether the target action is an overt bodily action or a mental action such as a decision (Vihvelin 2013: 111).

Vihvelin says that when subject to such intervention Jones is unquestionably in a bad way (Vihvelin 2013: 97). But, she says, Jones still has all he needs for free will. This is because he is at least able to *try* to do otherwise. Vihvelin treats being able to try otherwise as a significant ability, and this must be allowed if Black is a Bodyguard. In such cases Black uses the beginnings of Jones's *action* to know when to intervene. If, therefore, it is stated that Black's intervention is triggered by Jones's trying to decide otherwise, it must be allowed that this trying to do otherwise is an action on Jones's part (Vihvelin 2013: 97). Moreover, once we allow this, it is easy to see how possessing this ability is relevant to free will. Jones has two options available to him: he is able to act in the way that Black wants 'on his own' and he is able to try to decide otherwise. The fact that he's able to try to decide otherwise means that what Jones does do is up to him – it doesn't matter one bit that Jones's action in the so-called 'alternative sequence' is cut short.

I agree with Vihvelin when she says that when Jones is subject to this kind of intervention he still has free will with respect to the action he did perform. However, Vihvelin's account of why this is so is badly mistaken. First there is a tension in Vihvelin's position: she allows that trying is an action, and that Jones can try to decide otherwise. But she also maintains that Jones is unable to perform any action of a different type to that which he does perform. But if trying is an action, then if Jones had tried to decide otherwise, he would have performed an action of a different type. It doesn't matter that his trying would not have been successful. Second, Jones isn't able to try to decide otherwise in virtue of the ability that Vihvelin ascribes to him. Vihvelin attributes to Jones a particular, non-maximal ability, namely, the particular ability to try to decide otherwise in the relevant test-cases. But if the conclusion of 3.3 is correct, then the abilities relevant to free will are particular, maximally specific abilities. Now the point is not that Jones doesn't have the ability Vihvelin ascribes him; nor is the point that Jones isn't in a relevant test case – he is. The point, rather, is that because this ability is fairly general, it is not the end of the story concerning whether Jones can indeed decide otherwise.

Now consider the case where Black is a Pre-emptor. When Black is a Pre-emptor Vihvelin says that he removes *none* of Jones's abilities. All of Jones's narrow abilities and, more surprisingly, all of his wide abilities, remain intact when Black the Pre-emptor is around (Vihvelin 2013: 107). Vihvelin wants to maintain this while also maintaining that Black's presence as a Pre-emptor makes it the case (a) that Jones *will never try to* perform an action of a different type and (b) that Jones *will never* perform an action of a different kind. In other words, given a Frankfurt-style case where Black is a Pre-emptor, it is impossible for Jones to try to perform a different action and it's impossible for Jones to succeed in performing a different action. But this relative impossibility – impossibility given Pre-emptor Black's presence – tells us nothing about any of Jones's abilities, according to Vihvelin. She summarises her assessment like this: 'even if Black has the power to predict and alter everything and anything that Jones thinks or does, this power does nothing to diminish Jones's ability to choose or act otherwise' (Vihvelin 2013: 101).

Vihvelin is able to make this assessment because of the way she ties abilities to a set of relevant test cases. (I argued that such a strategy cannot be employed without first illicitly assuming something akin to what I have called the definitional circumstances, but here I am ignoring that worry). Assuming that the conclusions drawn in 3.3 are correct, it should be obvious where Vihvelin's account goes wrong: she makes her assessment based on abilities that are not relevant to free will. However, it is instructive to provide an assessment of these cases using the account of abilities so far developed.

I will use one of Vihvelin's Frankfurt-style cases. The case is called **BIKE** and is as follows. Jones lives in a deterministic universe, and he faces a choice between going for a walk and going on a bike ride. Pre-emptor Black wants to stop Jones going on a bike ride.[9] To that end, Black monitors Jones waiting for the reliable 'prior sign' which will indicate what Jones will do before he even begins to do it. This means if Jones would, apart from Black's intervention, decide to go on a bike ride, Black would detect this and manipulate Jones's brain so that he's unsuccessful. We are to consider a particular occasion, however, where Jones decides to go for a walk 'on his own'; Black's intervention was not necessary. Given this scenario, Vihvelin says that:

> Jones could have ... gone for a ride on his bike. He had the ability; he knows how to ride a bike, and the relevant parts of his brain and body were functioning correctly – no broken limbs, loss of muscle control, pathological fear of bike-riding, and so on. He also had the opportunity; his bike was right there [and] in good working order (Vihvelin 2008: 356).

This passage makes it clear that when assessing whether Jones could have done otherwise, Vihvelin addresses three questions corresponding to the three kinds of ability she recognises. She asks whether Jones has the skill ('he knows how to ride a bike'), she asks whether Jones has the narrow ability ('his brain is functioning correctly....he has no pathological fear of bike-riding') and she asks whether Jones has the wide ability ('he also had the opportunity'). On Vihvelin's view, there is just one ability in each category to ask about: *the* skill, *the* narrow ability, *the* wide ability. Jones has each of these abilities, so Vihvelin concludes that Jones is able to go on bike ride – in any sense you please – and so had free will with respect to the decision he made.

It's important to be clear that both the narrow ability and the wide ability that Vihvelin asks about are genuine abilities, and Jones (presumably) does have both of them. The problem, once more, is that this is only part of the story. There are many other narrow and wide abilities (in my terminology: non-particular and particular abilities) to go for a bike ride that are obscured from Vihvelin's view. For example, we could ask about Jones's ability to go for a bike ride in 40 degree C heat, his ability to go for a bike ride through London traffic, and of course, most pertinent to the current discussion, his ability to go for a bike ride given Pre-emptor Black's presence.

Does Jones have the ability to go for a bike ride given Pre-emptor Black's presence? The answer to this question is no. To possess an ability an agent needs to succeed in performing the action (were they to try) in at least some of the range of possible scenarios demarcated by the definitional circumstances. The more cases the

[9] Vihvelin refuses to call this case a Frankfurt-style case because Black is aiming to stop Jones doing something, rather than aiming to ensure Jones does something; I do not think this difference matters.

agent succeeds in, the 'stronger' the agent's ability will be. But given Black's presence Jones succeeds in no cases. So Jones has no abilities to go for a bike ride which include Black's presence as part of the definitional circumstances. This means that Jones does not have the ability-to-go-for-a-bike-ride-given-Pre-emptor-Black's-presence. This is not a maximally specific ability, of course; but given that Jones succeeds in zero cases where Black is present, Jones will not have any abilities to go for a bike ride given Black's presence, regardless of their specificity. As a result, we can conclude that Jones is not able to decide to go for a bike ride in any sense relevant to free will.

That does not quite settle things, however. I have agreed that possessing the ability to try to do otherwise is enough for free will (more on this in chapter 5). But in that case it might be that although Black's presence rules out Jones's going for a bike ride, it doesn't rule out Jones's trying to go on a bike ride. If that is the case, then Jones might yet have the ability to try to do otherwise even given Black's presence. The way the case is constructed, though, makes it clear that this is not the case. Given Black's presence, Jones's trying is also impossible: Jones won't succeed in trying to decide otherwise in any of the possible scenarios which include Black as a Pre-emptor. And so we can attribute to Jones neither the ability-to-try-to-go-for-a-bike-ride-given-Black's-presence nor the ability-to-try-to-decide-to-go-for-a-bike-ride-given-Black's-presence. And again, although these abilities are not maximally specific, since Jones succeeds in zero of these cases, we can conclude that Jones will also lack all of the abilities more specific than these, including the maximally specific abilities. Jones, therefore, lacks free will with respect to his decision about whether to go for a bike ride.

The foregoing discussion accepted the common assumption – made by Vihvelin and Whittle, among others – that an ability to A may be construed as a disposition to A in response to the agent's trying to A. The final points just made, however, help to expose a lacuna in accounts which accept this assumption. If abilities are treated as dispositions to perform an action in response to the agent's trying to perform the action, it's not at all clear how we are to understand the ability to try. We have seen above that Vihvelin treats trying as an action, and so accepts that abilities to try are sometimes relevant to free will – they are sometimes relevant to an agent's control. This, I think, is correct: trying is an action and when an agent can try to do something but cannot succeed in that endeavor, that is still provides the agent with a significant option. But to make sense of this we either need to drop the assumption that abilities can be treated as dispositions where trying features as the stimulus conditions, or we need to provide a separate account of what it is to be able to try. Given that trying is itself an action, there seems little reason to think that abilities to try should be of a different form to all other abilities. This makes the first option preferable, and in the next chapter I will consider an account which pursues this line of thought.

Chapter 4 – Possible worlds based analyses of 'can'

4.1. Lehrer's '"Can" in theory and practice'

4.1.1. Lehrer's motivation and aims

Lehrer was a prominent critic of the conditional accounts of 'can' and in chapters 1 and 2 I concluded that his counter-example to such accounts, as interpreted by Chisholm, was decisive and refutes even contemporary conditional accounts. In its place, Lehrer (1976; 1990) offered what he called a 'possible worlds analysis of "can"'. Lehrer's target is, of course, the sense of 'can' relevant to free will. This is the sense we appeal to when providing excuses for actions and Lehrer uses this thought to give a rough idea about what this sense of 'can' involves:

> I wish to fix on a sense of ['can' and 'could have'] that will always suffice to defeat an excuse for failing to perform an action on the grounds of incapability. To ensure that a person does not get off the hook on the grounds of incapability, we must be able to claim that she could have performed the action, in a very strong sense of 'could have'. It is a sense implying that all those conditions required for the person to succeed in performing the action were fulfilled. If the action requires a special skill, the person has it; if instruments are needed, she has access to them; if information is essential, she is informed, and so forth (Lehrer 1990: 44).

In this passage Lehrer is saying that if the agent can perform some action in the relevant sense – the sense to be analysed – then it will not be possible to excuse the agent on the basis of inability or incapability. This requires a 'strong sense' of 'can' because, and this point has been made frequently in the preceding chapters, sometimes an agent is able to A in one (or more) sense(s) and yet is not able to A in another sense. The sense Lehrer wants to fix on is 'strong' inasmuch as it precludes this possibility: if the agent could have A-ed in Lehrer's sense, then there can be no sense in which the agent was unable to perform the action. Lehrer identifies this sense of 'can' with Austin's 'all-in' sense but he is keen to note that, contrary to what some authors have claimed, there is more to this sense of 'can' than ability plus opportunity (Lehrer 1990: 44). This claim depends on how one understands both ability and opportunity, and as is evident from the preceding chapters, 'able' and 'ability' vary in meaning almost as much as 'can'. Lehrer provides an example, however, which indicates that he has in mind the skill sense of 'ability'. Consider a man who has the ability to jump seven feet. If that man is terrified of heights then he will be unable to jump a six-foot chasm that lies before him (Lehrer 1990: 44). The man envisaged here has the skill to make the jump, and he has the opportunity, but he lacks various psychological states necessary for performing the jump. Lehrer is thus using the term 'ability' in a way which excludes such states of mind from consideration, but he's clear that he thinks the sense of 'can' relevant to free will needs to include them.

The sense of 'can' which Lehrer aims to capture is thus closely related to what Vihvelin attempted to capture with her notion of wide ability. As we saw in the chapter 2, on Vihvelin's view, an agent has the wide ability to A if the agent has the relevant skills, the relevant psychological states, and the relevant opportunities. Lehrer and Vihvelin, then, are in fairly close agreement as to the phenomenon that we should be focused on if our interest is the sense of 'can' or 'able' most relevant to free will.

According to Lehrer, this sense of 'can', which is used in moral, social and political contexts, is in tension with what he calls the '"can" of theory' (Lehrer 1976: 241). The '"can" of theory' is the sense of 'can' associated with scientific and rational inquiry. Such theoretical inquiry gives us good reason to think that human behaviour and action is governed by the laws of nature and entirely explicable in terms of those laws and the antecedent conditions (Lehrer 1990: 46). This is not to say that such explanations are currently available to us, just that there are such explanations. Lehrer, writing in 1976, said that it was reasonable to assume that the scientific explanations of all human actions are deterministic. Whether Lehrer did not consider the indeterministic nature of physics well-confirmed at the time, or whether he had in mind the idea that any indeterminism at the fundamental level would be cancelled out at the macro-level, is not clear. Lehrer describes the tension between the 'can' of practice and the 'can' of scientific inquiry as follows: if there are such deterministic explanations, and if the antecedent conditions which feature in them are likewise deterministically explained by antecedent conditions and the laws, then every human action is 'determined ancestrally' (Lehrer 1990: 47). What this means is that every human action is determined by the laws and a set of conditions which obtained before there were any humans. And this suggests that there was no time at which any agent could have done other than she did do. Suppose Kate faces a choice between chocolate and vanilla and chooses vanilla. Lehrer's point is that although we can very easily *imagine* situations where Kate chooses chocolate, *from the theoretical point of view* the implication is that Kate could not have done anything different (Lehrer 1990: 47). This is clearly in conflict with the way 'can' is used in the moral realm. Lehrer's intention is to resolve this tension in the way we use 'can' 'in theory and practice' (Lehrer 1976: 241).

As mentioned, Lehrer's analysis uses the framework of possible worlds. And it is worth being clear the role this framework plays. In calling his analysis a 'possible worlds analysis' (Lehrer 1976), Lehrer gives the impression that some substantial work is done by the framework itself, perhaps that there is a class of analyses which, in virtue of being expressed in terms of possible worlds, have something in common, something which may even go some way to addressing the question whether free will is compatible with determinism. As we will see in more detail in section 4.3.2 this is not the case. The possible worlds framework does allow for analyses of 'can' to be formulated which could not be formulated in terms of conditionals (Lehrer's own analysis is one of these).

But if, as perhaps most philosophers think, something like Lewis's semantics for counterfactuals is correct, then conditional analyses of 'can' may be also expressed using the possible worlds framework.

To label an analysis a 'possible worlds analysis', therefore, is not particularly informative. Lehrer's analysis would be better classed as a possibility based analysis. The key difference between such accounts and conditional accounts is as follows. All conditional accounts say that an agent can A because *if* such-and-such a condition *were* to be the case, then the agent would (or would probably) perform action A. Moore employed the *agent's choosing to A* as the antecedent; Vihvelin employed the *agent's trying to A*. But whatever antecedent is chosen, conditional accounts determine whether the agent is able to do A by first assuming that the antecedent is satisfied and then asking what happens given that assumption. Some conditional accounts – e.g. Moore's and Vihvelin's (2004) account – require that the agent ***would definitely*** A, assuming the antecedent is satisfied. This means that the modal connection between the agent's trying (choosing) and the performance of the action is one of restricted necessity: there is a set of salient possible worlds where the agent tries to A, and in order to count as being able to A, the agent must A in all worlds in that set.

Vihvelin's (2013) account retained the conditional structure but weakened the modal connection between antecedent and consequent. As we saw in chapter 2, Vihvelin, employing insights from Manley & Wasserman's account of dispositions, said that an agent is able to A if and only if, were the agent to try to A, the agent would A in a 'suitable proportion of cases'. Here, the agent doesn't need to A in all the possible trying-to-A cases in order to count as being able to A; rather, the agent just needs to A in a 'suitable proportion' of such cases. The important point is that this account, while an improvement, still makes being able to A dependent on what happens *given the assumption that a certain non-actual event or state obtains*.

It is precisely this feature of conditional accounts which gives rise to what I called the **Transfer Problem**: if the antecedent cited by a conditional account is not possible for the agent, then the action is not possible for the agent. Lehrer was one of the first to push this complaint and his account differs because it drops this conditional structure entirely. So instead of saying (roughly) that S is able to A if and only if, *given a certain assumption B*, S would (probably) A, Lehrer says that S is able to A if and only if there is a minimally different possible world, satisfying some minimal conditions, where S A-s. This produces the following distinctive feature: the minimally different world might differ from the actual world only at times *after* the attribution of the 'can' statements. This opens up the possibility that Lehrer's analysis might be successful because the world is indeterministic; or, to put it differently, Lehrer's analysis is not inherently compatibilist.

4.1.2. The initial desiderata for the analysis

In section 4.1.1 I outlined the phenomenon that Lehrer takes himself to be analysing: a sense of 'can' which takes into account the agent's skill, mental or psychological states, and opportunities. The first thing to say about this sense of 'can' is that it is an indicative sense of 'can', and as such the analysis will apply both to the present tense 'can' and the past tense 'could'. But here we need to forestall a confusion. As Austin taught us (see chapter 1) 'could' is used in two very different ways. Sometimes it functions as the past indicative of 'can' and sometimes it occurs in the subjunctive mood, in which case it might be either in the past or present tense. When used as a past indicative, 'could' either means something like 'I used to be able to do things of this type' or something like 'I was in a position to perform that action', depending on whether it is used with the 'have' modifier. For example, in the statement 'In 1980 I could take the bus across town for 50p, and on the 28th of July 1980 I could have taken the 2:02pm bus across town for 50p', both uses of 'could' are past indicative, with the first referring to something that I could in general do and the second referring to a particular occasion when I could have performed an action. It is this kind of past indicative use of 'could' that Lehrer's account applies to; moreover, as he includes the 'have' modifier it is clear that he's articulating the truth conditions for statements which claim that the agent could have performed a particular action rather than statements which claim that the agent was able to perform actions of a certain type. This makes sense: when used without the 'have' modifier 'could' means something like 'I used to be able to perform actions of a certain type' and this is naturally taken as equivalent to the attribution of a skill. But as we've seen, Lehrer's target sense of 'can' concerns more than the attribution of a skill.

A general feature of such 'can' statements, something that Lehrer has previously argued for (Lehrer and Taylor 1965), is that they have a double time-index. A statement such as 'S can A' is more properly written as 'S can (at t_1) do A (at t_2)'. The first time index is the time at which the agent can perform the action; the second is the time at which the action is performed. Two indices are needed because these times can differ. Lehrer provides an example to illustrate this: suppose a man borrows some money (at t_1) with the promise to pay it back a week later (t_7). Suppose at the time of incurring the debt, he has enough money to repay the debt a week later: he can at t_1 repay the debt at t_7. But he might mishandle his finances the day after incurring the debt (t_2) such that he can no longer repay the debt on the promised day: he cannot at t_2 repay the debt at t_7 (Lehrer 1990: 45).

As mentioned, Lehrer's analysis of 'can' and 'could' statements is framed in terms of possible worlds. Before stating his account, however, Lehrer uses that same framework to make the problem itself more precise. He does this by first providing the definition for a determining condition, which is as follows (Lehrer 1990: 48). Consider an action A performed at t_n in the actual world W. Then:

Condition C occurring at t_i in the actual world W determines A at t_n in W ($i < n$) if and only if:

(1) there is a possible world in which C occurs at t_i and A does not occur at t_n, but
(2) for any possible world w having the same laws as the actual world W, if C occurs at t_i in w, then A occurs at t_n in w.

Lehrer explains as follows. Clause (1) ensures that it is logically possible for C to occur and A not to occur. This precludes C logically implying A. Clause (2) ensures that it is nomically necessary that if C occurs A will also occur. Lehrer then says that:

> Action A at t_n in W is determined if and only if there is some condition in W that determines that A occurs at t_n in W.

So far this does not allow us to capture the incompatibilist's worry. Indeed, depending on how this is read, it might be that most actions are determined according to the above definition even on an incompatibilist's understanding of action. For example, if time t_n refers to the time at which the action is completed, then all those actions which are performed by first performing another action might turn out to be determined on this definition: my raising my arm (at t_n), for example, might be determined by my deciding to raise my arm (at t_{n-1}). On the other hand, if we understood action A to be a stretch of activity, and understood t_n to be a time interval, then an action's being determined would be more problematic. What this shows is that much depends on the conception of action one employs when reading the above definition. Although Lehrer isn't explicit, everything he says implies that A stands for the result or outcome of an action and t_n refers to an instant. Lehrer makes explicit the consequence of the thesis of determinism by defining ancestral determination as follows (Lehrer 1990: 48):

> Action A in W is ancestrally determined if and only if
> (1) there is a condition C occurring at t_i ($i < n$) in W that determines A at t_n in W and
> (2) for any condition C occurring at any time t_h ($h < n$) in W such that C at t_h in W determines A at t_n in W, there is a condition B occurring at t_g ($g < h$) in W that determines C at t_h in W.

The idea is intuitive enough: an action A is ancestrally determined if and only if it is determined by condition C, and that condition C is itself determined, and so on, for each determining condition. With this consequence of the thesis of determinism stated, we can specify a further condition that any compatibilist account of 'can' must satisfy. Such an account must show how statements of the following form are consistent (Lehrer 1976: 245):

> (1) S did not do A at t_2
> (2) S's not doing A at t_2 was ancestrally determined
> (3) S could (at t_1) have done A at t_2

Notice that this formulation of the problem makes use of the idea that 'can' statements have a double time index. We are now in a position to summarise the three initial desiderata that Lehrer thinks any compatibilist analysis of 'can' must satisfy: (a) it must be an analysis of that sense of 'can' which takes into account the agent's abilities (skills), the agent's mental states, and the agent's opportunities, (b) it must recognise and accommodate the double-time index on 'can' statements, and (c) it must be such that the truth conditions for the analysis judge that at least some agents could, on at least some occasions, have done otherwise than they actually did even given the assumption that what they did was ancestrally determined according to the definition above.

4.1.3. Lehrer's analysis

Lehrer takes the notion of a possible world to be primitive but intends his account to be compatible with any of the dominant views about the metaphysical nature of possible worlds (Lehrer 1990: 47–8). Within such a framework, to say that S can (at t_1) A (at t_2) is to say that there is a possible world where S performs action A. By restricting our focus to different sets of possible worlds, we arrive at different kinds of possibility. If we put no restrictions on the possible worlds to be considered then we employ a notion of logical possibility: in this case, 'S can A' would be true if there is some possible world where S A-s (and any possible world will do). If we assess the statement 'S can A' by considering only those possible worlds which have the same laws of nature as the actual world then we would have a notion of nomic possibility (Lehrer 1990: 48). Given a set of criteria which define a notion of possibility, the worlds matching those criteria are said to be *accessible*. In providing a possible worlds account of 'can' the challenge is to specify in detail the restrictions that should be put on the set of possible worlds; in other words, the challenge is to say which worlds should count as accessible.

The first suggestion Lehrer considers derives from his critique of the conditional account. Recall from section 1.2.3 that the primary failing of the conditional analyses was that the conditional used in the analysans may be true even when the antecedent cited in the conditional cannot (in some significant sense) be true. Lehrer's **Red Candy** example illustrated this: the conditional 'if Lehrer choose to take a red candy, he would' is true, but Lehrer has a pathology which precludes him choosing to take a red candy, so the statement 'Lehrer chose to take a red candy' cannot be true. And because he cannot *choose* to take a red candy, he cannot *take* a red candy, regardless of the truth of the conditional.

Lehrer says that this observation might lead us to think that 'those cases in which the conditional statements are true and yet the person could not have performed the action are ones in which some necessary condition for the performance of the action is missing' and from this thought we might suggest the following analysis of 'could' (Lehrer 1990: 59):

> S could have done A at t_n only if no necessary condition for doing A at t_n was lacking.

Rewritten in terms of possible worlds:

> S could have done A at t_n at the actual world W only if there is a possible world w where S does A and where every condition at w which was necessary for S's A-ing also obtained at W.

Such a view is inadequate, however. Lehrer uses the following example to show why:

> Imagine that I leave the fingers of my left hand relaxed at t_n. From the simple fact that I do this, it would be peculiar to suppose that I could not have clenched my fingers into a fist instead. Yet, there is a certain muscle in my arm, *flexor digitorum profundus* to be precise, that must be flexed for my hand to be so clenched, and that muscle is, in fact, unflexed (Lehrer 1990: 59).

Lehrer's suggestion here is that the flexing of the *flexor digitorum profundus* muscle is a necessary condition for his performing the action of clenching his fingers into a fist, and that condition is unsatisfied. The current proposal, therefore, would judge that Lehrer could not have clenched his fist, which is absurd. Intuitively, the lack of this necessary condition is benign because the agent can do something about it: if Lehrer were to clench his fist the necessary condition would obtain.

What we need is an analysis which does not require that *every* necessary condition for the performance of the action obtain, but which imposes more restrictions on the accessibility relation than bare logical possibility. Lehrer identifies three criteria which form the basis of his restriction on the relevant possible worlds, and which therefore together determine whether a possible world is accessible or not. First, the possible world must have the same laws of nature as the actual world (Lehrer 1990: 59). There are possible worlds with different laws of nature where Lehrer clenches his fist but they 'hardly show that he could have performed that action' (Lehrer 1990: 60). Second, the accessible possible worlds must be allowed to differ from the actual world, in order to facilitate the absence of some necessary conditions, but they must only differ minimally. Worlds which 'differ wildly' from the actual world, even if they have the same laws of nature, 'are irrelevant to the question of what a person could have done in the actual world' (Lehrer 1990: 60). There are, for example, possible worlds 'wildly different' from the actual world where (a) I can perform actions that I cannot perform in the actual world (e.g. there is a world where I can play Rachmaninoff's second Concerto because I learned the piano from a young age and continued in daily, diligent practice) and (b) I cannot perform actions that I can perform in the actual world (there is a world where I cannot raise my arm due to a childhood accident that resulted in permanent paralysis). These worlds are irrelevant to what I can do in the actual world.

These first two conditions are still not restrictive enough, however. Lehrer provides another example to show why: suppose that Will is chained to a wall from which he wants to move away. Will doesn't have the means to rid himself of the chains. Nevertheless, we can find a possible world with the same laws of nature and which only differs minimally from the actual world where Will does move away: for example, a possible world where the chain is broken. Again, however, 'it hardly follows' from this that Will could have moved from the wall (Lehrer 1990: 60).

The reason, Lehrer says, is that the minimal change we've made – i.e. the breaking of the chains – bestows upon Will an *advantage* for his moving away from the wall. Lehrer does not define 'advantage'. Instead, he provides an intuitive characterisation by contrasting the above case with one where, intuitively, there is no advantage present: suppose Will is stood up against a wall and is not restricted by any chains. He does not move away. Here 'it is natural to assume that [Will] could have moved away from the wall' and that his doing so required no advantage which he lacks in actuality (Lehrer 1990: 61). Combining these three restrictions on the possible worlds which should count as accessible we arrive at the following analysis:

> (**Analysis 1**) 'S could (at t_i) have done A at t_n' is true in W **if and only if** there is a possible world w having the same laws as the actual world W and only minimally different from W so that 'S does A at t_n' is true in w in such a way that S has no advantage at t_i for doing A at t_n in w that he lacks in W, and t_n is past in W.

This is straightforward enough: each of the three criteria outlined above has become a restriction, and together they dictate which possible worlds are accessible and so which worlds will ground either the truth or falsity of the 'could' statements we're interested in. We avoid the problem with the first suggestion – i.e. the idea that S could have A-ed (where A is unperformed) only if all conditions necessary for A-ing were present – by allowing the possible worlds to differ, but only minimally, from the actual world. We then tighten this requirement by stipulating that the minimally different possible worlds must not bestow any advantage upon the agent. In other words, when considering whether an agent could have A-ed, we consider only a subset of the minimally different possible worlds.

This account is close to Lehrer's final position, but there is still a problem. **Analysis 1** successfully excludes those worlds where the agent possesses, at t_i, an advantage for the performance of A at t_n (where t_i is before t_n) such as, for example, the advantage Will possesses for moving away from the wall in the possible world where the chains are broken. But this restriction isn't exhaustive enough because sometimes a possible world bestows upon the agent an advantage for performing the action but not at t_i: the advantage is bestowed at some time between t_i and t_n. Lehrer (1990: 62) provides the following example to illustrate this kind of case:

> Suppose that Dan has promised to repay a debt in a pawnbroker's shop at a specific time, say between 1 and 2 o'clock on Friday. On Thursday, Dan has the money and is not far from the appointed place. But on Friday morning a tornado strikes the pawnbroker's shop, sucking it up like a vacuum, and, as a result, Dan is unable to repay the debt at that place at the appointed time. ... Dan could not have repaid the debt ... as promised. For Dan promised to repay the debt at the pawnbroker's shop on Friday between 1 and 2 o'clock, and that he could not have done because of the tornado.

The point of this example is as follows: Dan cannot repay the debt *as promised*; Lehrer thinks this is our intuitive judgement about the case, and one that bears up under reflection. But there is a possible world which satisfies the criteria laid out in **Analysis 1** and yet is such that Dan does repay the debt. For example, there is a possible world, call it w, where the tornado takes a slightly different path and so avoids hitting the shop. The tornado only strikes on Friday, which means that *on Thursday* Dan in the actual world, W, has all the same advantages as Dan in w, and this is all that **Analysis 1** requires for the truth of the relevant 'could' statement. Dan in w does have, *on Friday morning*, an advantage that Dan in W does not have. But **Analysis 1** doesn't range over times; it stipulates only that the agent must not possess any advantages *at* t_i (the time of the attribution of the 'could' statement). It therefore gives the result that Dan could (on Thursday) have repaid the debt (on Friday between 1 and 2 o'clock) even though the tornado is going to destroy the pawnbroker's shop on Friday morning.

It might be thought that what is needed here is simply to allow the exclusion of advantages to range over all times between t_i (the attribution of the 'can' statement) and t_n (the time when the action is performed). But things are not so simple. Suppose that Adina is at the airport waiting in the boarding area to get on a flight that leaves in fifteen minutes. Lehrer says that Adina's being at the boarding area fifteen minutes before takeoff is an advantage she possesses for getting the flight, an advantage that Adina wouldn't possess if, for example, she had decided to stay at home. As with the *flexor digitorum profundus* example, however, we should not think that had Adina stayed at home (such that she'd then lack the advantage of being at the airport in time), worlds where she is at the airport need to be excluded from consideration when judging what Adina could have done. Again, this is because it is natural to think that Adina is in control of that advantage: there is something she can do – e.g. get a taxi to the airport – such that if she did it she would obtain the advantage of being at the airport. Moreover, doing that thing (i.e. getting a taxi) requires no further advantage, just as Lehrer's clenching his fist required no further advantage.

To summarise: we start with the set of minimally different worlds. The example with Will chained to the wall suggests that some minimally different worlds – those bestowing advantages – need to be excluded from consideration. The Dan in debt example shows that the worlds which should be excluded might bestow advantages on the agent at any time between the ascription of the 'can' statement (t_i) and the performance of the action (t_n). But we cannot introduce a blanket ban on advantages: the Adina at the airport case shows that some worlds bestowing advantages should not be excluded from consideration on those grounds, intuitively, because the advantage was a result of something the agent did. This forces Lehrer to introduce a distinction between advantages: those that should be excluded and those that should not. The former he calls *inadmissible* advantages and their possession render possible worlds inaccessible to the agent. The latter are called *admissible* advantages and the possession of such advantages does not render the world inaccessible.

These ideas come together in the following analysis (Lehrer 1976: 256):

> (**Analysis 2**) 'S could (at t_i) have done A at t_n' is true in the actual world W **if and only if** there is a possible world w having the same laws as W and minimally different from W so that 'S does A at t_n' is true

in w in such a way that any advantage S has in w for doing A at t_n which he lacks in W is admissible for S from W and t_n is past.

An advantage S lacks in W is admissible for S from W **if and only if** either:

(a) the advantage results from S doing something B at t_j ($t_i \leq t_j \leq t_n$) when he has no additional advantage for doing B at t_j in w which he lacks in W, or:

(b) the advantage results from S doing something C at t_k ($t_i < t_k \leq t_n$) when S has no additional advantages for doing C at t_k in w which he lacks in W except those advantages admissible to S from W resulting from what S does prior to t_k.

In **Analysis 2** there is no time index attached to the possession of the advantage. So an advantage possessed at any time between t_i and t_n potentially excludes the possible world from being accessible. But this is then qualified with the introduction of the notion of an admissible advantage: when an advantage is admissible it does not preclude the possible world counting as accessible. Advantages are admissible if they result from something the agent does. The notion of *doing* employed here is to be understood broadly: decisions, choices, tryings are all to be included in what an agent does (Lehrer 1990: 64). The idea of an advantage 'stemming from' something the agent does is to be understood recursively: Adina's being in the airport lounge at, say, t_5 is an advantage which results from her getting through security with ample time to spare. Her being at the entrance of the airport ready to go through security is an advantage she has at, say, t_4, and is itself a result of her getting into a taxi at t_3. Adina's getting a taxi is an advantage resulting from her booking the taxi at t_2, and so on. The advantage Adina possesses at t_5 – her standing at the entrance to the airport – is admissible only because it results from a sequence of actions which she herself performed. There is a possible world where Adina is stood in front of the entrance at t_4, not as the result of a taxi ride she instigated, but because a mystery chauffer inexplicably arrived at her house, forced her into the car, and transported her to the airport in time for her flight. In such a world, Adina's being at the airport entrance would not be an admissible advantage but would instead be inadmissible. The notion of admissibility, therefore, is a historical one: it depends on the relationship between the advantage and the agent's past actions.

Lehrer's analysis yields one immediate benefit over conditional accounts: it applies to all of the agent's actions. In section 1.2.3 we saw that conditional analyses, which typically employ the agent's choosing to A or trying to A in the antecedent of the conditional, struggled to explain the sense in which the agent was able to choose or try. Vihvelin's dispositional account also had this structure and suffered from the same problem. Such analyses and accounts had to maintain that the sense in which an agent can or is able to choose (to decide, to try) is different from the sense in which the agent can perform all other actions. Aune attempted to get round the problem by saying that choices were not actions (see 1.2.3); Vihvelin skirted the issue entirely (see 2.2). For Lehrer this problem simply doesn't arise. He attempts to analyse the relevant 'could' statements directly by describing when an action – any action – is possible for an agent. His account thus unequivocally applies to decisions, choices and tryings.

How does Lehrer's analysis solve the problem he set out to address, as he put it, the problem of reconciling the 'can' of theory and practice? Lehrer described the problem as a matter of showing the following statements to be consistent (Lehrer 1976: 245) (see 4.1.2):

(1) S did not do A at t_2
(2) S's not doing A at t_2 was ancestrally determined
(3) S could (at t_1) have done A at t_2

Suppose then, that S does not do A at t_2 and his not doing so was ancestrally determined. Does Lehrer's analysis of 'could have' allow us to affirm (3)? That depends, Lehrer says, 'on whether the presence of an antecedent condition determining the person's inaction entails the lack of some advantage he needed to perform the action' (Lehrer 1990: 72). If the presence of a condition which determines that S does not A entails that S lacks an advantage for the performance of A then any possible world (with the same laws of nature) where S does A will be one in which S possesses an advantage not possessed in the actual world (because the condition determining the non-action is missing).

Lehrer argues, however, that conditions which determine that the agent does not A do not entail the lack of an advantage for the performance of A. To do this he employs the *flexor digitorum profundus* example once more:

> [S]uppose I do not clench my fingers into a fist at a specific moment. That *flexor digitorum profundus* [the muscle in the forearm which flexes the fingers] was unflexed just prior to that moment determines the fingers not being clenched. Yet it hardly follows from the antecedent condition of that muscle being unflexed that I could not have clenched the fingers. On the contrary, I could have clenched the fingers, and had I chosen to do so, *flexor digitorum profundus* would have been in a flexed state at the required time. The non-occurrence of an action at a time being determined and ancestrally determined by antecedent conditions does not entail that the person in question would require some advantage he lacks in the actual world for him to perform the action (Lehrer 1976: 263–4).

But, says Lehrer, 'it hardly follows from the antecedent condition of that muscle being unflexed that I could not have clenched the fingers' (Lehrer 1976: 264). That would be an absurd conclusion. Instead, it should be clear that in the above scenario Lehrer could have clenched his fist. On the basis of this example Lehrer concludes that 'the non-occurrence of an action at a time being determined and ancestrally determined by antecedent conditions does not entail that the person in question would require some advantage he lacks in the actual world for him to perform the action' (Lehrer 1990: 72–3). That conclusion, we will see shortly, is far too hasty. In particular, the example provides no support for thinking that conditions which *ancestrally* determine the action performed fail to entail the lack of an advantage. I will argue for this at length in section 4.3.2, after first outlining a more general problem with Lehrer's account in 4.3.1. Section 4.2 considers and rejects one kind of counter-example brought against Lehrer's analysis.

4.2. Horgan's donation counter-example

A number of writers have raised putative counter-examples to Lehrer's account. Here is a paraphrase of Terence Horgan's (1977) example which is perhaps the most pressing:

> Jones has $2000 in the bank earmarked for a family holiday. When his aunt rings to ask for a $2000 donation to a dog charity she's involved with, Jones declines to give the money, despite being generally keen to help his aunt whenever he can. What Jones didn't know, however, was that he'd just received a large inheritance ($100,000, say) from a long lost relative. Moreover, earlier in the day the executor of the estate tried to ring let him know his good fortune. Unfortunately, the line was busy and so Jones remained unaware of his inheritance (Horgan 1977: 404–5).

Horgan's assessment of this example is as follows. First, it is part of the design of the case that Jones had earmarked the money for his family holiday and was intent on using the money for that purpose. His psychology and motivational structure was such that if Jones had donated the holiday money to his aunt this would be understood as significantly out of character. So although we are told that Jones is generally sympathetic to his aunt and her requests, he was opposed to giving her the money on this occasion. Still, it was not psychologically impossible for Jones to give the money to his aunt. That is, we are to suppose that given Jones's psychology, and assuming he has no knowledge of his inheritance, it was possible, even if extremely unlikely, that he would have given the money to his aunt. Put another way: it is stipulated that Jones was such that, if we imagined the same scenario but without the inheritance, he was able to donate the money to his aunt.

So Jones, aside from any considerations to do with the inheritance, was able to donate the money. And clearly, the addition of the inheritance should not affect this. Not only does Jones not know about the inheritance, but if he were to know about it, it should (if anything) make it easier for Jones to donate the money. Thus, Horgan thinks that what we want to say about this example is that Jones could have donated the money (i.e. he was able to donate the money).

Horgan contends, however, that Lehrer's analysis gives the wrong result. There are two possible scenarios where Jones donates the money which might be pertinent to the truth of the relevant 'could' statement. First, Jones might have simply decided to give his aunt the money, despite having no knowledge of his inheritance. In this scenario, Jones would be acting against the reasons he has as well as his settled conviction to use the money for his family holiday. As Horgan tells it, there would be a 'significant change in [Jones's] dominant wants in a possible world where he decides to sacrifice his family's holiday in order to donate the money – to wit, he suddenly wants to donate the money more than he wants his family to have their holiday' (1977: 406). Still, it is possible that Jones does this: it's part of the description of the case that it is not psychologically impossible for Jones to choose in this way. In this possible world Jones's donating the money does not require any inadmissible advantages; that is, it does not require any advantages which have not resulted from anything

Jones himself has done. Indeed, it does not seem as if Jones requires *any* advantages to perform the action in this possible world. So this world is a candidate for making the 'could' claim true. Call this possible world 'w1'.

There is another possible scenario where Jones donates the money, namely, one where Jones successfully receives the phone call about his inheritance. In this scenario, Jones possesses an advantage for donating the money which he did not possess in the actual world, namely, knowledge of his inheritance. Moreover, this advantage is inadmissible because it results from something that was outside of his control. That is, in this possible scenario, he receives the good news in virtue of his phone line not being busy, which is not down to him. To forestall one objection: it might be suggested that Jones *is* in control of whether the line is busy, and so is (in some sense) in control of whether he receives the call. Perhaps Jones could have finished up the previous call quicker thus leaving the line free. This objection is misguided, not just because the sense in which Jones 'controls' whether the call from the executor of his long lost relative's estate gets through is not the relevant sense (I will argue for this claim in the following chapter), but because the example is easily altered such that it does not involve Jones acting at all. We can simply suppose that in the actual world the line was busy because someone else tried to ring Jones a split second before the executor rang. The executor gets the engaged tone, and there is nothing Jones could have done that would have prevented that. We can then suppose that in the possible world now under consideration the first caller rings *two seconds later* such that the executor's call gets through instead. Call the possible world where this happens 'w2'.

Horgan's key idea is that w2 requires less of a departure from the actual world than does w1. World w1, where Jones gives the money on his own, would require a 'significant change in Jones's dominant wants', whereas world w2, where the executor's call gets through, requires only a small change, namely, the first caller rings two seconds later (Horgan 1977: 406). The problem, Horgan thinks, is that because w2 requires less of a departure from the actual world, according to Lehrer's analysis it will be the world which settles the truth value of the statement 'Jones could have donated the money'. And because Jones has an inadmissible advantage in this world, Lehrer's analysis will give the result that Jones was not able to donate the money, which is contrary to the description of the case.

I will argue that this example does draw attention to a problem with Lehrer's analysis, but the problem is not what Horgan takes it to be. To see this it is helpful to distinguish two subtly different ways of understanding Horgan's complaint.

One way of construing the case is as follows. Each of the possible worlds where Jones donates the money (w1 and w2) is *minimally different* from the actual world. In w1 Jones has no advantages for donating the money while in w2 Jones has an inadmissible advantage. World w2 requires less of a departure from the actual world, so it is the one which determines the truth value of the 'could' statement, and in virtue of the inadmissible advantage Jones possesses in w2, Lehrer's analysis thus gives the result that Jones cannot donate the money. Thus, Lehrer's analysis gives the wrong result.

This way of construing the counter-example fails. First, Lehrer's analysis does not say that the truth of the 'could' statement is settled by whichever minimally different world is *closest* to the actual world. It requires only that there is *a* possible world minimally different such that Jones donates the money without relying on any inadmissible advantages. And on this understanding of the example there is such a world, namely, w1.

There is a second way of understanding the counter-example, however, and it avoids the problem raised above. On this reading of the example, we have the two worlds where Jones donates the money, w1 and w2, but it is not the case that both count as minimally different. The world where Jones donates in virtue of a 'significant change in his dominant wants' (w1), it might be suggested, requires such a big divergence from the actual world that it does not count as being minimally different from the actual world. If it does not count as minimally different, however, it is precluded from view as far as the assessment of the 'could' statement goes. The world where Jones donates in virtue of an inadmissible advantage (w2), however, only requires a minor departure from the actual world and so it does count as being minimally different. This latter world is therefore the one that determines the truth value of the 'could' claim and because of the inadmissible advantage in w2, that claim comes out false. Thus, Lehrer's analysis gives the wrong result.

It is not immediately clear which reading Horgan is employing. At one point he says that 'a possible world in which Jones donates the money, but which otherwise differs minimally from the actual world, will not be a world where his motives are so radically altered that he decides to sacrifice his family's holiday' (Horgan 1977: 405). This seems to suggest that Horgan thinks none of the minimally different worlds will be ones where Jones's motivates have the radical changes needed if Jones is to donate the money without any inadmissible advantages; this suggests that Horgan intended the second reading. On the other hand, Horgan explains that 'the counter-example rests on the fact that [Jones's] change in motives is substantially greater than the change consisting of Jones' telephone not being busy when the crucial call comes through. The former change is the greater one because of the strong intensity of his (actual-world) motives' (Horgan 1977: 406). This might suggest the first reading.

In fact, both understandings of the example are flawed inasmuch as they presuppose that that the two possible worlds, w1 and w2, may be compared to see which one departs more from the actual world. But this is an assumption that Lehrer is committed to rejecting given his use of the notion of *minimal difference*. Why? The notion of minimal difference comes from Pollock's possible world semantics for counterfactuals. It operates as a selection function which takes as arguments a proposition, P, and a possible world, A, and gives as a result that possible world which is minimally changed from A such that P is true and all necessary adjustments are made for consistency (if there are multiple ways A can be adjusted to accommodate P then it returns a set of possible worlds). The notion of minimal difference thus plays a role in Pollock's semantics of counterfactuals similar to that played by the notion of *comparative similarity* in the Stalnaker/Lewis semantics.

One key feature of comparative similarity is the assumption that any two worlds may be compared to see which one is *closer to* the actual world. Pollock rejects this assumption. I will not recount Pollock's argument against comparative similarity – he argues in depth that it leaves certain principles invalid which should be valid (Pollock 1976: 18–22). It is enough to say that Pollock thinks the solution is to reject the idea that any two possible worlds may be compared to see which one is closer to the actual world. Pollock does not think this is the metric we use when assessing counterfactuals. Instead we employ a simpler rule which Pollock's notion of *minimal difference* aims to capture. The basic idea is that in assessing a counterfactual of the form 'P □⟶ Q' we need to assess whether Q is true in the worlds which are *minimally changed from the actual world so that P is true* (and the relevant changes have been made to ensure consistency). It's clear that Lehrer employs the notion of minimal difference because he too considers it mistaken to assume that any two possible worlds may be compared against a third (Lehrer 1990: 51–2).

Given that both readings of the Horgan's example rely on the idea that w2 is much closer to the actual world than w1, both readings are flawed. Indeed, the second reading is flawed twice over. First, it is illegitimate to ask which of w1 or w2 is closer to the actual world on the basis of Lehrer's analysis alone. As his analysis employs the notion of minimal difference there is simply no way to answer this question. We can of course ask the question by appealing to the Stalnaker/Lewis notion of comparative similarity; but then the answer will be irrelevant to Lehrer's analysis of the 'could' statements under consideration. This point alone dooms Horgan's entire approach to finding a counter-example. Horgan's approach is to construct a case where features of the actual world that (intuitively) should not affect the modal claims being assessed nevertheless do *if those modal claims are analysed using counterfactuals*, because those features affect the ordering of worlds with respect to comparative similarity. But Lehrer's analysis does not employ comparative similarity, so this approach has no traction.

Second, it does not make sense to ask whether w1 – the world where Jones gives the money despite not having received news of his inheritance – is 'too far away' to count as minimally different. Possible worlds are not minimally different to the actual world *simpliciter*. They are minimally different with respect to some proposition. As mentioned, the notion of minimal difference is defined as a function which takes a world and a contrary to fact proposition and returns a set of worlds which are 'minimally changed' so as to make that proposition true. So if P is the contrary to fact proposition under consideration, the only possible worlds where P is true that do not count as minimally different are those where there are other arbitrary changes that are not necessitated simply by ensuring consistency. Thus, it doesn't matter how big of a departure from the actual world it requires to envisage Jones deciding – on his own and out of character – to donate his holiday money to his aunt, as long as we don't envisage any other unrelated changes, then *by definition* such a possible world will be minimally different. So Horgan's counter-example fails, however it is understood.

Still, it does highlight something important about Lehrer's analysis, namely, that the notion of minimal difference is not doing much work at all – certainly nothing like the work done by notions of comparative similarity in accounts which employ a Lewisian counterfactual to analyse or provide an account of the sense of 'could' or 'able' relevant to free will (e.g. Vihvelin's account, considered in chapter 2). After all, minimal difference only adjusts for logical consistency. What does this mean for Lehrer's analysis? Suppose we alter the example above in the following way: Jones doesn't have $2000 in his bank, he's never inherited anything, and he hates his aunt and never gives her anything – indeed, he's never donated a thing to charity in his life. To top it all off, Jones has recently died. His aunt, unaware of Jones's demise, rings to ask him for the money. Clearly, it's impossible in a number of senses for Jones to donate the money: it's 'biologically impossible', given Jones's death, it's 'financially impossible', given his lack of finances, and on this revised scenario it would have been (had Jones been alive) 'psychologically impossible' for Jones to donate the money. Still, there is a possible world minimally different from this new scenario such that Jones decides to donate the money. In this world Jones has not died, has a radically different character, and has lots of money. The minimal difference function allows us to change as many things as possible in order to arrive at world where Jones donates the money.

Clearly, such a possible world should not be pertinent to the truth value of the relevant statement concerning what Jones could do in the original scenario. And on Lehrer's analysis it isn't. But this is not because of anything to do with the notion of minimal difference per se, it is rather because of the proposition Lehrer passes to the minimal difference function. Recall that worlds minimally different from the actual world are always minimally different *such that some contrary to fact proposition is true*. And the contrary to fact proposition Lehrer passes to the minimal difference function is not simply *that Jones donates the money* but rather *that Jones donates the money in a way that does not rely on any advantages he does not have in the actual world*. This is why neither the example immediately above – where Jones is dead – nor either reading of Horgan's putative counter-example falsifies Lehrer's analysis. What this does illustrate, however, is the crucial role that the notion of advantages play in Lehrer's analysis; in the following section I will consider and endorse an objection which claims that Lehrer's understanding of admissible and inadmissible advantages is crucial, inasmuch as they rely on a prior understanding of the agent's abilities.

4.3. Objection: the notion of an advantage depends on that of ability

4.3.1. Circularity worries

The notion of an agent possessing an advantage for the performance of some action is a central feature of Lehrer's analysis. He introduces the idea with a number of examples: in the possible world where Will is unchained, or where Will's chains are broken, Will possesses an advantage he doesn't have in the actual world where he is chained. And in the possible world where the tornado takes a slightly different route and misses the pawnbroker's shop, Dan possesses an advantage that he doesn't possess in the actual world. But no analysis or account is ever given of the notion of an advantage, and so we have to assume that, like the notion of a possible world, it is a primitive of Lehrer's analysis.

Two early critics, Horgan (1977) and Fischer (1979), have argued that this is problematic because, in fact, we cannot understand the notion of an advantage without first understanding what it is for an agent to possess an ability and/or an opportunity. Lehrer is clear that the sense of 'can' he is analysing is not exhausted by the notions of ability and opportunity. The sense of 'can' he's focused on also takes into account the agent's mental state, for example. Nevertheless, the sense of 'can' under investigation does include the agent's possessing various abilities and the agent's possessing various opportunities. If, therefore, this sense of 'can' is analysed in terms of advantages, but advantages in their turn cannot be explained without appealing to the ideas of ability and opportunity (among other things), then Lehrer's analysis is conceptually circular.

Horgan puts the worry like this: an agent's having an advantage 'is most naturally understood in terms of ability and opportunity' (Horgan 1977: 407). An advantage for doing A, Horgan continues, is a condition or state which enhances or increases the agent's ability or opportunity for A. It might be that an agent, S, already has the ability to A and the opportunity to A, in which case if S gains an advantage for the performance of A, A-ing will in some way become easier: S's ability is increased, or S gains a further opportunity. Alternatively, if S doesn't have the ability or opportunity to A, then gaining an advantage for A will produce the ability or opportunity to A.

These points seem plausible and make good sense of Lehrer's examples. In the possible world where Will's chains are broken, we judge that he has an advantage not possessed in the actual world because he has the opportunity to exercise some of his abilities, an opportunity he doesn't have in the actual world. In Lehrer's famous **Red Candy** example (see 1.2.3), we take him to have an advantage in the possible world where he chooses to take a candy, and this is plausibly thought to be because he has an ability in that world which he lacks in the actual world, namely, the ability to choose to take a red candy (his pathology in the actual world has destroyed that ability).

Fischer puts the worry slightly differently. He says the following:

> Lehrer has not offered a useful, 'reductive' analysis of 'ability'. By a reductive analysis, I mean the analysis of a disputed, unclear, or controversial notion in terms of less disputed, less vague, or less controversial notions. The problem is that Lehrer's analysis makes crucial use of the notion of 'having an advantage'; Lehrer never provides an account of 'having an advantage in a possible world which one lacks in the actual world', and it is hard to see how Lehrer could provide an account of 'having an advantage' without making reference to the notion of 'ability'. And if we take 'having an advantage' as primitive or unanalyzed, then it is unclear whether our intuitions are any less confused and disputed about 'having an advantage' than about 'ability' (Fischer 1979: 53).

Here, Fischer is using a broad sense of 'ability' which corresponds roughly to Lehrer's use of 'could' – that is, Fischer is using a sense of 'ability' which refers to the analysandum. This contrasts with Horgan's use of 'ability' above which is evidently restricted to a narrower sense of 'ability', one which does not include the notion of possessing an opportunity. (This nicely illustrates one difficulty mentioned at the outset of the current investigation, namely, that the terms 'able' and 'ability' have almost as many meanings as 'can' and 'could' and that it is therefore very important to be clear about which meaning is intended, especially when one concept is being explicated in terms of the other.) Fischer, employing this broad sense of 'able', goes on to say that one natural explanation of the sentence 'I have an advantage in possible world p toward doing A which I lack in the actual world' is that I am able in possible world p, but not in the actual world, to do something that would result in A. Here we have the same idea as that found in Horgan: the most natural way to understand an agent's possessing an advantage is in terms of a person's being able to do something. The point, I take it, isn't *just* that the notion of an advantage is 'most naturally' understood in terms of ability. It's also that there is no feasible alternative. If the notion of advantage were most naturally understood in terms of ability, but there were another explication which was almost as natural, and which invoked no ability-related concepts, then Lehrer's account would not be in as bad a position as Horgan and Fischer take it to be. Their criticism relies on the claim that it's just not clear how else the notion of an advantage could be explicated.

Of course, that the objection relies on this further claim doesn't necessarily weaken it. It is, after all, hard to see how else the notion of advantage could be explicated. For example, at one point Lehrer describes the effect of possessing an advantage in terms of prevention and obstacles. Would it help to appeal to such concepts? Lehrer says that if, in the actual world, a person is prevented from doing A, then there will be no possible worlds meeting Lehrer's criteria where the person does A (Lehrer 1990: 65). This is because if someone is prevented from doing A, then there is an obstacle in the way of her doing A. This shows that in the minimally different possible worlds where the person does do A, and that obstacle is thus removed, Lehrer takes the removal of such an obstacle to be an advantage. So perhaps we can explicate advantages in terms of prevention and obstacles.

But, as Fischer (1979: 53–4) convincingly argues, the notions of *prevention* and *obstacle* are concepts which form part of a closely interrelated set of notions along with *ability* and *having an advantage*. All of these concepts are subject to dispute, and a person's intuitions or judgements about whether an agent is prevented from doing something will tend to correspond appropriately to his or her intuitions or judgements about whether the agent can or is able to do that thing. For example, someone who is convinced by incompatibilist arguments to the effect that determinism renders agents unable to do otherwise, will also think that if determinism is true, the conditions which determine that some particular action will take place, will be conditions that prevent – which are an obstacle to – any other action taking place. Indeed, that might be just what incompatibilists mean when they affirm that, if determinism is true, the agent is unable to do otherwise. I think this point is correct and as a result consider the circularity objection problematic for Lehrer. In the following section I will outline what follows from this objection.

4.3.2. A consequence of the circularity objection

It is important to be clear about what the circularity objection shows if it is, as I take it to be, correct. To begin with, one thing it *doesn't* show is that Lehrer's analysis is inconsistent. Moreover, it doesn't show that Lehrer hasn't accurately captured the relationship between concepts like ability, opportunity and having an advantage. What it does show, however, is that Lehrer's analysis, as it stands, is incapable of delivering an answer to the question of whether free will is compatible with the thesis of determinism, which was Lehrer's stated aim. In Lehrer's terms, his account is incapable of demonstrating the reconciliation between an agent's being ancestrally determined and the agent's being able to do otherwise.

We can see this by considering the following question: given that S does A in the actual world at t_n, and assuming that there is, at time t_i, an ancestrally determining condition for S's doing A at t_n ($i < n$), should the lack of such that ancestrally determining condition in possible world w where S does not A count as an advantage for S's not doing A? If so, then Lehrer's account does not reconcile determinism and free will. Lehrer gives no argument to answer no to this question: he gives no reason to think that ancestrally determining conditions won't count as advantages. Obviously, if they count as advantages they will be inadmissible advantages, as such conditions obtain before the agent exists and so cannot be the result of something the agent does. What Lehrer does say about ancestrally determining conditions comes in the context of a discussion of his *flexor digitorum* example. Recall that the example was as follows:

> [S]uppose I do not clench my fingers into a fist at a specific moment. That *flexor digitorum profundus* [the muscle in the forearm which flexes the fingers] was unflexed just prior to that moment determines the fingers not being clenched. Yet it hardly follows from the antecedent condition of that muscle being unflexed that I could not have clenched the fingers. On the contrary, I could have clenched the fingers, and had I chosen to do so, *flexor digitorum profundus* would have been in a flexed state at the required time. The non-occurrence of an action at a time being determined and ancestrally determined by

> antecedent conditions does not entail that the person in question would require some advantage he lacks in the actual world for him to perform the action (Lehrer 1976: 263–4).

As I noted in section 4.1.3, Lehrer concludes from this example not just that there are *some* actions which have *some* determining conditions and yet are such that the agent could have done otherwise, but that the agent's being able to do otherwise is compatible with *all* of the agent's actions being *ancestrally* determined. But this stronger conclusion does not follow from the example. This is easy to see if we make the time indices explicit: let t_3 be the 'specific moment' of which Lehrer speaks when he does not clench his fingers. Let t_2 be a moment 'just prior to' t_3; in that case, at t_2 the *flexor digitorum profundus* in Lehrer's forearm will not be flexed and the muscle's being in this state will determine his fingers not being clenched into a fist at t_3. We can make moment t_2 as close as we need to t_3 in order to secure this result. That is, t_2 can be *after* the latest time at which Lehrer could have made a decision that would've resulted in his fist's being clenched at t_3. Call the latest time at which Lehrer could have made such a decision time t_1. Lehrer's muscle being unflexed at t_2 is sufficient for, and so determines, the state of Lehrer's fist being unclenched at t_3 because at t_2 it's too late for him to do anything about it: if he were to decide or try to clench his fist at t_2, his fist's clenching would not occur until some point after t_3.

Note two things. First, Lehrer's fist being relaxed at t_2 *only* determines the state of Lehrer's fist (it's being unclenched) *at t_3*. It *doesn't* determine the state of his fist at any other time. Second, suppose that Lehrer actively refrained from clenching his fist. Then it is important to recognise that *the state of* Lehrer's muscle at t_2 determines only *the state of* his fist at t_3. It doesn't determine *his action* of leaving his fist unclenched, i.e., his refraining. His refraining is an action which is temporally extended and proceeds, on a coarse-grained view of action individuation, from t_1 until t_3. The start of his action is thus before the occurrence of the condition which determines his muscle's being unflexed. This makes it evident that the state that Lehrer identifies cannot possibly determine his *action* – and it is his action with which we are most concerned. What Lehrer's example shows, therefore, is only that it is possible for the end result of some action to be determined by a condition which occurs in between the agent's starting to act and the time at which the end result comes to obtain. He has not provided an example where (a) an agent performs an action which is ancestrally determined and (b) the agent could have done otherwise.

One interesting consequence of this is that because the disagreement is primarily over what counts as an advantage (or what counts as an admissible advantage), and not over the structure of Lehrer's analysis, incompatibilists could in fact adopt Lehrer's general framework and approach to the truth conditions of the relevant 'can' statements. Both Horgan (1977: 409ff) and Fischer (1979: 60) comment on this. If the incompatibilist adopted such an account then they would argue in one of two ways depending on whether they were, for sake of argument, accepting Lehrer's assumption that determinism is true. That is, assuming determinism is true, incompatibilists could argue that Lehrer's own account gives the right (from their point of view) result: suppose that S does A at t_1. Given determinism, there was a condition C which ancestrally determined that S would do A. And in any possible world S does not do A, there was a condition D which ancestrally determined that S would not do A. Now, the incompatibilist might ask the following: what could be more advantageous than a condition which ancestrally determines the non-performance of A? And they might suggest the correct answer is nothing. After all, such a condition is an 'absolutely fundamental, rock-bottom, metaphysical advantage' for the non-performance of A (Horgan 1977: 410). So in every possible world where S does A, S has an advantage for doing so. Of course, Lehrer would disagree, but the disagreement would be over what counts as an advantage and not over how to specify the truth conditions for the relevant 'can' statement. Alternatively, if we adopt Lehrer's analysis of 'can' but do not assume determinism to be true, then the relevant 'can' statements will be true just when the accessible worlds are causally accessible, which again, the incompatibilist will argue is the right result.

4.4. Objection: the need for an epistemic criterion on world accessibility

Lehrer's account of the 'can' relevant to free will fails to show that an agent's being able to do otherwise is compatible with determinism and as argued in the previous section, both compatibilists and incompatibilists could make use of the analysis. The disagreement between the compatibilist and the incompatibilist becomes, if we adopt Lehrer's analysis, a disagreement over the nature of advantages, which is something Lehrer left unanalysed. Nevertheless, Lehrer's analysis is a big improvement over the structure of conditional accounts such as Moore's and Vihvelin's. Despite this big leap forward, however, there is reason to think that as it stands Lehrer's analysis is too permissive: it includes too much in what an agent is able to do. Consider the following scenario:[10]

> (**Trapped Tom**) Tom is in a building which is currently ablaze and, having more knowledge of the building than those he is with, he is attempting to lead them to safety. With many passages blocked by the fire, there is only one route available to the exit. Tom is aware of the route, but halfway to the exit the group encounters a door protected by an electronic keypad lock. Tom does not know the combination and, moreover, was under no obligation to know the combination for the lock.

Intuitively, Tom cannot lead those he's with to safety: not knowing the combination, he cannot or is unable to open the door. Given that fact, it seems reasonable to think that Tom could not be blamed for the deaths of those he's with. Lehrer's analysis, however, delivers a different verdict. It would judge that Tom could have opened the lock and thus led the people out of the building. The reason is that Lehrer's analysis judges an agent

[10] This example is a simplified version of one from Peter Morriss (1987: 55); similar scenarios also appear in Graham Oddie and Pavel Tichý (1982).

is able to do anything that can be done by performing a series or sequence of actions which the agent knows how to do. This is because Lehrer puts no restrictions on what can take the place of 'A' in his analysis.

There is a sequence of actions, each of which Tom can do, and which would result in the door being unlocked. For example, suppose that the combination to the lock is 4353. Tom can punch each of those digits into the keypad, and he can punch each in after the other. If he did so, the door would unlock. Because Tom knows how to punch a single digit into a keypad, it appears that Lehrer's account will have to judge that Tom could have unlocked the door: there is a possible world where this happens.

One initial response to this might be to say that such a possible world – one where Tom punches in the correct sequence and so unlocks the door – would not ground the truth of the 'can' statement on Lehrer's view because that world would be one where Tom has an advantage that he does not have in the actual world, namely, knowledge of the combination to the lock. But this is false. Tom doesn't need such knowledge to punch in the correct sequence. There is a possible world where Tom punches in the right sequence of numbers without any knowledge that it is the right sequence. For example, a world where Tom just guesses or punches in a random set of digits might be one where he gets lucky. Tom needs no advantages to punch each number in, so he needs no advantages to punch the whole sequence in. Therefore, according to Lehrer's analysis, Tom could have unlocked the door. Although this is a problem for Lehrer, it is not fatal. The solution is to add a knowledge condition somewhere in the account. In the following chapter I will explore a number of different senses of 'able' which require different epistemic conditions to be satisfied.

4.5. Campbell's 'A theory of compatibilist alternatives'

Joseph Keim Campbell (1997) defends a compatibilist understanding of the ability to do otherwise using a possible worlds framework similar to the one put forward by Lehrer. He is defending classical compatibilism against the challenge posed by Frankfurt-style cases and not that posed by incompatibilists based on the problem of ancestral determination. As a result his immediate aim is to show that the following is false:

> (**Semi-compatibilism**) Moral responsibility does not require alternative possibilities of action, so neither does any freedom that is necessary for moral responsibility (Campbell 1997: 319)

Campbell begins with a possible worlds framework, and although he does not explicitly identify any problems in Lehrer's account, he treats Lehrer's account as incomplete. Campbell thinks that if the task is to explicate the sense of 'can' relevant to responsibility, then once we assume a possible worlds framework, that task equates to articulating the details of an accessibility relation which can be used to model the sense of 'can' in question (Campbell 1997: 323). Campbell's first three conditions on the accessibility relation are very similar to the conditions imposed by Lehrer (although Lehrer doesn't frame things in terms of the accessibility relation) (Campbell 1997: 324):

> (**C1**) Accessible worlds must have the same laws of nature as the actual world.
> (**C2**) In accessible worlds, agents cannot have any abilities or capacities that they lack in the actual world.
> (**C3**) Any advantage that a person has in an accessible world which they lack in the actual world must result from something that the person does in that possible world.

Condition **C1** is a repeating of Lehrer's position on the laws of nature. Campbell notes that some – in particular those with a Humean conception of laws – might dispute this and require that the laws be allowed to vary slightly. He puts that issue to one side and I will do the same. Condition **C2** makes explicit some of the motivation behind Lehrer's introduction of advantages: some worlds should be excluded because they bestow illegitimate benefits or advantages on an agent, and Campbell identifies abilities and capacities as paradigm examples of such advantages. Caution is needed here, however. Neither Lehrer nor Campbell say much about what they mean by 'ability'. It was clear from Lehrer's discussion that he meant skill rather than any broader conception of abilities. The same is true of Campbell's account. **C2** should thus be read as prohibiting the possession of skills or competences that aren't possessed in the actual world. This reading is confirmed by the example Campbell gives in support of **C2**:

> Suppose that Eleanor is standing in a yard where a young kitten has accidentally fallen into a pool. … Unfortunately, Eleanor has not learned how to swim. Since she cannot swim, it is natural to say that she could not have saved the cat and [this] would suggest that she is not responsible for its death (Campbell 1997: 322, 324).

Eleanor's not being able to swim is a matter of her lacking a *skill* or *competence*. If this is what Campbell means, the position has some intuitive support, but it is not problem free: we saw in chapter 2 that circumstances play a prominent role in defining such properties. The phrase 'the ability to swim', it was argued there, does not pick out a single ability property, but rather a whole spectrum of such properties. To fully characterise such a property requires adding a set of circumstances to the definition. This was true of what I called general abilities, and it will also be true of skills. This will be a non-trivial condition for Campbell to fill out. Much more needs to be said concerning **C2**, then, if it is to prove informative. I will assume for the rest of this chapter that by 'ability' Campbell means skills and capacities to do things in those circumstances that would be considered ordinary.

Condition **C3** articulates a rule for the kind of advantages that are okay, namely, those which result from one of the agent's own actions. Campbell uses the following example to illustrate the importance of **C3**:

> Suppose that Eleanor is standing beside an outside pool when a cat falls into the pool. In order to save it, she need only bend down and pick it up. However, due to an unfortunate childhood incident, resulting in a pathological fear of cats, she cannot choose to pick the cat up and the cat drowns. Intuitively, she could not have saved the cat because of her neurosis (Campbell 1997: 324).

The point being made here is the one that was deemed decisive against conditional and dispositional accounts: although it is true that *if* Eleanor had chosen to reach down and pick up the cat, then she *would* have saved the cat, Eleanor has a neurosis which makes her unable to choose to do so. Because Eleanor couldn't choose, she couldn't have saved the cat. There are, of course, possible worlds where Eleanor does choose. In such worlds Eleanor has no neurosis. But we are after a sense of 'can' according to which Eleanor cannot save the cat, and so we need to exclude such worlds. **C3** is an attempt to articulate the reasoning which underlies our judgements in cases like Eleanor's.

There is reason to think that **C2** will result in an account that is too narrow. This is because there is a kind of ability – and one which appears to bestow a significant kind of control – that involves acquiring new skills during the course of some activity. With **C2** as a condition on 'able', however, Campbell's account will give the result that the agent is unable to do anything which requires the acquisition of some skill. To put the point differently, we need to allow that the advantages the agent is able to acquire through acting – those advantages described by Campbell's **C3** condition – include the acquisition of new skills. Here's an example to illustrate this point. Suppose that we need a Butterfly Knot tying. Among our party is Alice, an adept knot tier. Alice does not currently know how to tie a Butterfly Knot, but she can easily teach herself to do so using the knot tying manual that we have before us. No one else among our group can tie anything other than a simple bow, and no one else is any good at following written instructions concerning the tying of knots. There's a sense in which Alice *can* or *is able to* tie a Butterfly Knot. Put in these terms, this might not be immediately obvious: the terms 'can' and 'able' are most frequently used to discuss skills or competencies that an agent already possesses. But the sense at issue comes to the fore if we put it terms of bringing about a certain result: Alice can bring it about that a Butterfly Knot is tied; Alice can see to it that a Butterfly Knot is tied; and so on. Given the example just described, Alice is the only one who can bring such a thing about. Alice is in control over whether a Butterfly Knot is tied; moreover, Alice is in control over whether *she ties* a Butterfly Knot. To accommodate these kinds of cases, I will assume in what follows that **C2** should be qualified by **C3**; in other words, we should allow the acquisition of skills if they are acquired in virtue of something the agent does.

Campbell says condition **C3** might lead one to think that the conclusions which the semi-compatibilists draw from Frankfurt-style cases – namely, that the agent does not need to be able to do otherwise to be morally responsible, and so, whatever kind of free will is required for responsibility does not involve being able to do otherwise – are vindicated. Campbell presents a Frankfurt-style case in order to illustrate this:

> Suppose that a woman, Eleanor, and her father, Roscoe, decide to rob a bank since both are desperately in need of money. Despite her claims to the contrary, Roscoe fears that Eleanor may change her mind about the robbery at the last minute. As a fall-back, he has a device implanted in Eleanor's brain that, when activated, will render her unable to do anything other than follow through with the robbery as planned.

> As it happens, Eleanor is a willing subject and she performs the crime on her own, without the activation of the device. Eleanor is morally responsible for her action but, it seems, she could not have done otherwise (Campbell 1997: 320–1).

Why might **C3** lead one to agree with the semi-compatibilist's conclusions regarding the irrelevance of the ability to do otherwise to moral responsibility (and so any free will that being morally responsible requires)? Campbell explains as follows. Any possible world where Eleanor has the device in her brain is one where she robs the bank. The worlds where she doesn't rob the bank are worlds where there is no device. So she only avoids robbing the bank if there is no device. Those worlds – where she lacks the device – are the worlds that could ground the truth of 'Eleanor could have refrained from robbing the bank'. One of them must be accessible if we are to affirm that Eleanor could have avoided robbing the bank. But the lack of a device is an advantage that Eleanor has in each of those worlds. More importantly, it would, according to Lehrer, be an inadmissible advantage, and according to Campbell, fail condition **C3**. According to either of these accounts therefore, Eleanor could not have done otherwise. That means that if we share Frankfurt's intuitions we would have to conclude that responsibility (and so any free will or control required for responsibility) does not require alternative possibilities.

Campbell thinks this reasoning is premature. We were led to condition **C3** by a consideration of cases involving psychological compulsion such as the example with Eleanor and the drowning cat. But, Campbell says, such cases are not wholly analogous to Frankfurt-style cases. With cases of psychological compulsion the affliction suffered is a factor causally relevant to the subsequent behaviour. Eleanor's neurosis was one of the things which causally produced her refusal to pick up the cat. But in Frankfurt-style cases the implanted device plays no causal role. That is part of the definition of the cases: the device is there ready to be activated, but by stipulation it is only a backup, the victim goes ahead and performs the action 'on her own'. Campbell says that this difference suggests the following, final criterion:

> (**C4**) Accessible worlds need not include factors which are causally irrelevant to the performance of actions (Campbell 1997: 325).

This criterion is to be read as follows: accessible worlds need not include factors *present in the actual world* if they are causally irrelevant to the action performed in the actual world. Campbell says that 'the implanted device ensures that Eleanor robs the bank but it does not play any causal role. As such, worlds in which the device is absent do not give Eleanor any advantage' (Campbell 1997: 325). In other words, **C4** is to be taken as a partial fleshing out of what counts as an *advantage*. Both the incompatibilist defender of the Principle of Alternative Possibilities (PAP) and the semi-compatibilist opponent of the principle agree that the lack of device counts as an advantage, although of course they diverge on how to assess Frankfurt-style cases once that is accepted.

Campbell's idea is to undercut this common agreement by denying that the absence of a device counts as an advantage. This means that some of the worlds without the device are accessible and so Eleanor could have avoided robbing the bank. Campbell asks us to consider the following counterfactual in support of this idea:

> (**CF1**) If Eleanor had chosen not to rob the bank and the device had not been implanted, then Eleanor would not have robbed the bank.

Campbell goes on to say that 'counterfactuals like **CF1** are essential to confirming that Eleanor was causally responsible for the bank robbery' (1997: 326). If **CF1** is false or vacuous then Eleanor cannot be held responsible. This is because the falsity of **CF1** would show that Eleanor's rational causal powers were ineffective: if she chose not to rob the bank with no device present, but still didn't avoid robbing the bank, her ability to act on the basis of her choices should be questioned. But because **CF1** is true its 'antecedent entails the existence of a possible world in which Eleanor does otherwise' (Campbell 1997: 326). Campbell then asks whether that possible world is accessible to Eleanor and he concludes that there is no reason not to think so. We may suppose this alternative world has the same laws as the actual world and that Eleanor's rational causal powers are efficacious. The world lacks the Frankfurtian device, but this is where **C4** comes in: because the device in the actual world doesn't operate, the fact that it doesn't exist in the alternative world isn't an advantage. So the possible world which **CF1** describes is accessible and Eleanor possesses alternatives despite being in a Frankfurt-style case.

4.6. Campbell's C4 and the acting freely/freedom to do otherwise distinction

One potential problem with Campbell's **C4** criterion on the accessibility relation was presented in an early reply by Michael McKenna (1998). McKenna takes aim at Campbell's **C4** criterion by drawing on a distinction, initially made by Frankfurt (1971), between *acting freely* and the *freedom to do otherwise*. McKenna says that on Frankfurt's view, an agent *acts freely* if what they do issues from their own volitions and they are not subject to external influence whereas an agent has the *freedom to do otherwise* if they have the ability to adopt a different course of action to the one they undertake (1998: 260).

According to McKenna, what Campbell gets right is that the counterfactual he identifies is indeed, as he claims, relevant to whether Eleanor acts freely in her robbing of the bank (McKenna 1998: 261). It confirms, in other words, that Eleanor was 'causally responsible for the bank robbery' (Campbell 1997: 326). What it doesn't do is establish that Eleanor was *free to refrain from robbing the bank* – by which we can assume (given the wider context) that McKenna means *able to refrain from robbing the bank*. If these are two distinct things then the existence of certain possibilities, like the ones without the device being present, suffice to show that in the actual world Eleanor's behaviour issued from her own rational causal powers – i.e. that she *acted freely*. But those same possibilities do not establish that Eleanor has the *freedom to do otherwise* and so we can't conclude that Eleanor *could have* done otherwise. McKenna uses Locke's famous example to make his case. Locke describes a man taken into a room whilst sleeping. The door is then locked, but the man chooses to stay in the room. McKenna says that this case is parallel to Campbell's example. He asserts that the following counterfactual suffices to show that Locke's man exercised his own powers and capacities to stay in the room:

> (**CF2**) If the man had chosen to leave the room and the door had not been locked, then the man would have left the room.

In other words, Locke's man *acted freely*. But it doesn't follow from this that Locke's man had the *freedom to do otherwise*. This should be evident from the case: it is just not possible for Locke's man to leave the room, that is an alternative he doesn't possess. Campbell's example is the same. It demonstrates that Eleanor acted freely, but not that she has the freedom to do otherwise. McKenna thinks that this should be intuitively evident: the possible world Campbell appeals to is one where the device doesn't exist, and because there is nothing Eleanor can do to bring about that situation, such a world just isn't 'genuinely accessible' to her (McKenna 1998: 263 n.12). McKenna concludes that far from supporting his compatibilist notion of the ability to do otherwise, Campbell's criteria, if they accurately isolate the rational capacities and powers required for moral agency, in fact show that alternative possibilities are not needed for moral responsibility (McKenna 1998: 263).

There are two problems with McKenna's diagnosis of Campbell's account. The first is the reliance on the distinction between *acting freely* and the *freedom to do otherwise*. There is no doubt that one can distinguish between an action's flowing from an agent's volitional state and the idea of an agent's being able to choose and do otherwise. But both the classical compatibilist and the incompatibilist will resist the idea that if an action flows from an agent's volitional state, then the agent can be said to act freely. In the context of this debate, 'free will' and 'acting freely' are technical terms which need explication. In contrast to McKenna, the classical compatibilist and the incompatibilist will likely suggest that agents act freely whenever they exercise their free will (i.e. their freedom to do otherwise). They will contend, in other words, that the fact that an agent's action flows from the agent's volitional state does not settle the matter as to whether it was done freely. They will also suggest that it is illegitimate to use, as McKenna does, a distinction between acting freely and being able to do otherwise, in replying to Campbell. Typically semi-compatibilists draw the distinction between acting freely and being able to do otherwise only after they have accepted Frankfurt's argument; that is, it's the Frankfurt-style cases that motivate the distinction. Campbell has provided a suggestion for where those cases go wrong; proponents of the Frankfurt-style cases cannot therefore appeal to a distinction which derives from accepting the cases in order to respond to Campbell's reply.

A second problem surrounds the use that both Campbell and McKenna make of their counterfactuals. The problem, in short, is that these counterfactuals do not suffice to show that the agent acted freely in the actual sequence (even if we accept for sake of argument the distinction between acting freely and freedom to do otherwise). The reasons are precisely those which caused so many problems for the simple conditional analysis of 'can', but we can focus just on finks. Suppose McKenna's **CF2** is true:

> (**CF2**) If the man had chosen to leave the room and the door had not been locked, then the man would have left the room.

The truth of **CF2** does not suffice to establish that Locke's man stayed in the room freely. McKenna assumes that the truth of this counterfactual demonstrates that Locke's man's rational capacities are in working order, such that, given that he didn't try to leave the room, that must be because he didn't choose to. But Locke's man might have a finkish lack of ability: the locking of the door might have triggered some nefarious intervener to remove the man's rational capacities and his ordinary physical abilities such that if the door is locked he loses his general ability to make decisions and his ability to walk. In such a case **CF2** would be true but it would not be the case that in the actual world Locke's man stayed in the room as a result of his own volitions. The truth of the counterfactual, therefore, does not establish that Locke's man acted freely.

4.7. Campbell's C4 and causal relevance

Campbell's **C4** condition runs as follows:

> (**C4**) Accessible worlds need not include factors which are causally irrelevant to the performance of actions (Campbell 1997: 325).

This is plausible because it is based on observations about the causal origins of an action which might at first glance seem reasonable, namely, that sometimes facts about how an action is produced can preclude it being something for which the agent is responsible (Campbell thinks cases of psychological compulsion fit this bill). Suppose we accept this idea for the time being. One significant problem with Campbell's principle is that it simply misapplies this central insight. If facts about the causal history of an action can preclude it being something for which the agent is responsible, then when assessing whether some factor (e.g. the Frankfurtian device) is causally relevant we need to consider whether it is causally relevant to the action or behaviour which occurs *in the possible world under assessment*. Campbell claims that we can abstract away from the Frankfurtian device – and therefore that worlds without the device can be considered accessible – because the device isn't causally relevant to the action which is *actually performed*. But the device's being causally irrelevant to what Eleanor does in the actual world does not mean it's causally irrelevant to what Eleanor does (and is able to do) in other possible worlds.

If, therefore, we are trying to determine what is possible for Eleanor, and we want to use causal relevance as a guiding principle as to whether some factor (or lack thereof) should be considered an advantage, it is not enough to consider just the causal history of what actually happens. Some factor in the actual world may restrict the space of what is possible simply because it is potentially causally relevant. A different example helps to make this clearer. Suppose that Peter left his child – little Johnny – unattended for a short period in a fenced-in play area; Johnny did not go near the fence for the period he was unattended, and, we might suppose, Johnny (being very short sighted) was unaware of the fence. The fence was strong and there was no way for Johnny to get out. Later Peter's wife finds out that he left Johnny unattended and rebukes him with the following argument:

> I know that little Johnny played happily in the play-area for 30 minutes. But he could have run out onto the road, where he might've been killed! After all, the fence played no causal role in keeping Johnny where he was, so I'm entitled to abstract away from it when considering what Johnny could've done. There is a possible world where, ignoring the causally irrelevant fence, I see that Johnny runs out onto the road and is hit by a car. And you, Peter, are responsible for allowing such a situation.

Surely Peter, in defending the reasonableness of his behaviour, will be within rights to argue that what could have happened depends on the existence of the fence even though the fence was not, in fact, causally relevant. And so it is with Eleanor and the Frankfurtian device. To assess, using the 'causally relevant' criterion behind Campbell's **C4**, whether some factor (or lack thereof) counts as an advantage, we need to consider whether that factor is causally relevant in any of the possible worlds (which share our laws) where it exists. The lack of a Frankfurt-device would be an advantage on this view, because in many of the ways Eleanor's situation plays out the Frankfurt-device is causally relevant. For example, if Eleanor became adamant they'd made a mistake and insisted that she and her father hand themselves in right away, or if Eleanor wavered for days and finally decided to refrain, or if she got cold feet at the last minute – in each of these cases, and many more besides, the Frankfurt-device would be causally active. Once we repair **C4** according to the underlying motivation we will be unable to abstract away from the Frankfurt-device and will have to consider it an advantage, judging, contra Campbell, that Eleanor cannot avoid robbing the bank. This suffices to show that **C4** as it stands is an inadequate condition.

So far I have argued that *if* causal relevance is to be the criterion by which we judge whether some factor counts as an advantage, *then* it must be the factor's causal relevance to the action performed in each possible scenario in which the factor exists. This position can be supported by considering the reasons Campbell is led to causal relevance in the first place. Causal relevance is important, Campbell thinks, because it explains the difference between cases of psychological compulsion and Frankfurt-style cases: in the former '[the compulsion] is a causal factor in the production of the wrong action' whereas in the latter the device is causally irrelevant

(Campbell 1997: 325). This is why, when considering what, for example, a kleptomaniac can do, we must not abstract away from the kleptomania. If Kate is a kleptomaniac and as a result steals something from a shop, the kleptomania is causally relevant in the actual world, so we cannot consider worlds where Kate is not a kleptomaniac as accessible. Cases of psychological compulsion (such as kleptomania) and Frankfurt-style cases ensure that the agent does a particular thing, but according to Campbell the former serve as a excuse and the latter do not because the former are causally relevant in the actual world whereas the latter are not.

But Campbell's diagnosis of why ailments such as kleptomania serve as an excuse is incorrect. Cases of psychological compulsion serve as an excuse not simply because the ailment *is* causally efficacious but because the ailment *ensures* a certain outcome. Here's a counter-example to Campbell's idea that cases of psychological compulsion excuse only because the disorder is causally efficacious and not because the disorder ensures a particular kind of outcome:

> Kate suffers from resistible kleptomania. At 1pm, when Kate is at the store, various desires to steal bubble up within her. She recognises them for what they are and, at 1:05pm, is able to think about something else and able to continue dwelling on the desires. If she were to think about something else she would be able to leave the store without stealing anything. She doesn't exercise the ability to think about something else and instead pays close attention to the desires. Eventually they become so strong that they cause her to steal something from the store.

In this case Kate's kleptomania is causally efficacious but it does not serve as an excuse, or, at least, it does not completely remove any culpability on Kate's part. The degree to which Kate is responsible here is proportional to the degree to which she was able to do otherwise. If Kate has only recently become a kleptomaniac such that she doesn't quickly recognise the urges to steal, if those urges to steal are almost overwhelming, and if the period during which she was able to direct her attention elsewhere was fleeting, then Kate's ability to do otherwise was minimal and so her responsibility minimal. This is precisely because, in having only a minimal ability to do otherwise, her control is diminished. If, on the other hand, Kate has suffered with kleptomania for quite some time, is able to quickly discern when such urges to steal are arising, and has an extended period during which she is able to direct her attention elsewhere, then Kate's ability to do otherwise is more pronounced and so is her responsibility. The causal efficacy of the kleptomania could be the same in both cases, so Campbell is simply off target when he identifies that feature of cases of psychological compulsion as a pertinent difference.

Chapter 5 – Epistemic abilities

5.1. Introduction

So far we have considered various different aspects of the nature of ability properties. In chapters 1 and 4 it was noted in passing that there may be a sense of 'able' and 'ability' relevant to free will which has an epistemic element. This chapter takes up that topic, investigating the idea that the abilities relevant to free will are what Goldman (1970: 203) and Morriss (1987: 53–6) have called *epistemic abilities*.

The kind of case which lends most support to this idea is typified by the **Trapped Tom** example introduced in the previous chapter (see 4.4). Recall that Tom was trapped in a burning building with only one route of escape. That route was blocked by a locked door which could only be opened using an electronic keypad, the combination for which Tom does not know. The important point about that example was as follows: Tom, being a typical human adult, is able to punch any arbitrary series of digits into the keypad. Moreover, for any specified sequence, he can punch it in intentionally. That means that there is some sense in which he could punch in what is in fact the right combination. If he were to do so, he would find that the door would open and he would get to safety. Intuitively, however, there is a significant sense in which it does not seem correct to say that Tom is able to leave the building: it is not under his control, it's not something that is up to him.

I will argue that there are abilities which reflect Tom's lack of knowledge and that it is such abilities that are relevant to free will. To see this, we need to reflect a little on how free will is often discussed in the literature. Most philosophers consider free will important primarily because of its connection to morality. Van Inwagen, for example, says that it is part of the 'classical tradition' that he inhabits that free will is important because 'if there were no free will – if no one were able to act otherwise – then no state of affairs would be anyone's fault. No one would ever be morally accountable for anything' (van Inwagen 1989: 402). Here van Inwagen defines free will in terms of the ability to do otherwise. And it is implicit in his statement that possessing free will, being able to do otherwise, gives agents the kind of control over their actions that is required if those actions are to be imputable to the agents in a way which makes them morally accountable. The same underlying thought is found among philosophers who do not define free will in terms of the ability to do otherwise; indeed, many simply start by defining free will as the kind of control – whatever its nature turns out to be – that is needed for moral responsibility (e.g. John Martin Fischer and Mark Ravizza (1998); Derk Pereboom (2001: xiv); Kevin Timpe (2008: 12)).

One consequence of thinking of free will as the kind of control required for moral responsibility is that it becomes easy to think of free will as one among many distinct conditions, all of which need to be satisfied if the agent is to be morally responsible. Indeed, it is now orthodox to put free will as the control condition alongside, at the very least, an epistemic condition on being morally responsible. Sometimes further conditions are also suggested (Timpe (2008: 9-10,9 n.24), for example, mentions an authenticity condition). One potential pitfall with this approach is that the conditions tend to be treated in isolation from each other; this has the result that any necessary condition on moral responsibility which is epistemic in nature is often thought to be part of 'the epistemic condition on moral responsibility'. Such conditions are then considered irrelevant to the control that agents have over what they do. The main contention of this chapter is that this is a mistake: many of these epistemic conditions are relevant to the control the agents can exert.

I will begin in section 5.2 with a consideration of some of the examples used to motivate the need for a distinction between, on the one hand, a freedom or control condition on moral responsibility, and on the other, an epistemic condition on moral responsibility. I will then highlight a number of counter-intuitive implications of accepting this orthodox view. In section 5.3 I will argue that many of the epistemic conditions normally treated as part of a separate epistemic condition on responsibility affect what it is an agent is capable of controlling, and so should be seen as part of an account of free will. I will outline a number of different kinds of epistemic ability which reflect this change, and which bestow different kinds of control on agents. Finally, I will outline what I think are the minimal abilities required for free will.

5.2. The separation of the freedom and epistemic conditions on responsibility

As mentioned, the dominant approach treats free will as one of the conditions which needs to be satisfied if an agent is to be morally responsible. This freedom or control condition, as it is sometimes called, is contrasted with an epistemic condition on moral responsibility. These two conditions – the freedom condition and the epistemic condition – are sometimes traced back to a distinction found in Aristotle between two kinds of excusing condition: force and ignorance (although as Fischer and Ravizza (1998: 12 n.17) point out, it is unclear whether Aristotle had in mind anything very close to the modern concept of moral responsibility). The distinction is widely endorsed and the two parts are typically discussed separately. Fischer and Ravizza (1998: 13–4), for example, endorse the distinction before going on to give a book-length account of the freedom condition; prominent attempts at articulating the epistemic condition include those by Carl Ginet (2000) and Timpe (2008: 10ff). In this section I will review some of the examples thought to motivate the distinction. I then want to outline the full implications of this orthodox view, some of which, I will suggest, are highly counter-intuitive. I will endorse points made by Mele which show that even those who endorse a strong separation between freedom and epistemic conditions typically embed substantial epistemic conditions on the freedom condition. This should prompt us to ask whether some of those epistemic conditions treated separately would be better treated as part of the freedom or control condition.

One caveat is needed at this point: the issues considered here are typically discussed under the banner of 'the *epistemic* conditions on moral responsibility' or as part of accounts of *epistemic* abilities. However, as we will see in 5.3, the conditions articulated don't usually require that agents *know* what it is they are doing; often what's thought to be required is *belief* or some other psychological attitude. In this section I will follow the literature in using the phrase 'epistemic conditions' to refer to whichever psychological attitudes are required; in section 5.3 I will reserve that term for conditions which do require knowledge.

To begin, then, consider a pair of examples from Ginet. In the first example we have an agent who performs an action and is – purportedly – in control of that action but who is nevertheless not responsible for it:

> (**Simon**) Simon enters the hotel room he has just checked into and flips what appears to be, and what he takes to be, an ordinary light switch, but, to his surprise and consternation, the flipping of the switch sets off a loud fire alarm (Ginet 2000: 269).

Ginet's second example is meant to have the same structure but with respect to an omission:

> (**Herb**) Herb did not unlock the back door before leaving for work, and therefore later that day the plumber was unable to get in to repair the furnace and left a note saying that he would not be able to come again until next week. But Herb did not know that the plumber was scheduled to come that day—his wife made the appointment and forgot to tell him (Ginet 2000: 269).

The idea behind these examples is as follows. Simon performs the action of flipping the light switch and in so doing he also sets off the fire alarm. Setting off the fire alarm is something that resulted from an action Simon performed. As such, according to this orthodox view it is something he *controlled*. Simon controls the setting off of the alarm because he set off the alarm *by* flipping the switch and it is clearly the case that he was in control of the switch flipping. For Ginet, Simon possesses this control because he satisfies a 'could have done otherwise (CDO)' condition: it was open to Simon to have acted in a way that would have resulted in the fire alarm not going off. In other words, Ginet, who accepts the characterisation of free will in terms of the ability to do otherwise, is saying that Simon had free will with respect to his setting off the fire alarm. Still, Simon is not morally responsible for setting off the fire alarm. And this is because he *didn't know* it would result from his flipping the light switch (neither could he have been expected to know this). Simon's not knowing that his action would have this result is, on the current view, taken to be an additional condition on being morally responsible over and above his possessing free will. Similarly for Herb: he did not leave the door unlocked and in doing so he blocked the plumber's access to the house. This was an omission of his and so was something that was under his control, but again, he cannot be blamed for locking out the plumber as he *didn't realise* the plumber was coming. Herb is not blameworthy here because he was wholly unaware – again, through no fault of his own – of what he was doing.

Fischer and Ravizza present a similar case: Kit reverses his car and (unknowingly) kills his kitten. Fischer and Ravizza do not characterise free will in terms of the ability to do otherwise; they gloss it as the 'control required for moral responsibility' and go on to say that an agent exercises free will if their behaviour is produced by the right kind of mechanism – a mechanism that is sensitive to reasons. In any case, the result is the same: Kit's killing his kitten satisfies their account of free will, with the result that they are led to affirm that Kit freely kills his kitten, despite not being aware that he was killing his kitten.

How would this view assess the **Trapped Tom** example? Again there would be a separation of freedom and epistemic conditions. It would affirm that Tom is able to escape the burning building; this is something Tom can do because he is able to type in a sequence of digits which would, in fact, be the correct combination. This means that Tom has free will with respect to the action of escaping from the building. However, this view would hold that because Tom does not know what the correct combination is, Tom will not be morally responsible for failing to lead those he's with to safety. Presumably this view would explain any intuitions we might have which speak in favour of denying that Tom is able to escape the building by saying that they are rooted in our unwillingness to hold Tom morally responsible for failing to escape. This latter fact, so the view goes, should not be explained by appealing to a deficiency in Tom's abilities but rather to a deficiency in his knowledge.

The examples above are a representative sample of the kinds of case used to motivate the distinction. What are the implications of imposing a sharp separation between freedom and epistemic conditions? Consider again Ginet's Simon example: Simon enters a hotel room, flips what he (reasonably) takes to be a light switch, and sets off a fire alarm. Ginet is clear that Simon satisfies the control condition on moral responsibility; that is, Simon has free will. Like van Inwagen, Ginet takes free will to be about having the ability to do otherwise. He provides the following 'could have done otherwise' condition which an agent needs to satisfy in order to have free will with respect to bringing about H (I've simplified the presentation somewhat as much of the detail is not relevant to the present point):

(**CDO**) At some time t1 earlier than t2, either

(i) S acted in a certain way W such that it was open to S at t1 not to act in way W then, and acting in way W brought about H at t2, or

(ii) S did not at t1 act in a certain way W such that it was open to S to act in way W then, and acting in way W would have brought about H at t2 (Ginet 2000: 268).

Clause (i) covers the case where the agent acts in a certain way W – which brings about result H – and affirms that the agent could have refrained from acting in way W. Clause (ii) covers the case where the agent doesn't act in a certain way W – a way which would have brought about result H – and affirms that the agent could have so acted. My purpose in introducing Ginet's account of free will is to illustrate that for Ginet it is straightforward that Simon was in control of setting off the fire alarm: Simon acted in a way which brought it about that the fire alarm went off, and there was a time when it was open to Simon not to act in that way. The crucial point is that by affirming that Simon satisfies **CDO** with respect to setting off the fire alarm, Ginet is affirming that Simon exercised control over setting off the fire alarm. This means that on Ginet's account Simon's setting off the fire alarm was something he did freely: it was an exercise of his free will. This means, however, that Ginet thinks that Simon has the same level of control over setting off the fire alarm as he does over flipping the switch. Whatever kind of control is required in order to do something freely, Simon has that kind of control over both flipping the switch and setting off the fire alarm.

This last point, I submit, is counter-intuitive. If possessing free will is a matter of possessing the control required for moral responsibility, then it is implausible to say that Simon exercised his free will with respect to setting off the fire alarm – i.e. that Simon freely set off the fire alarm – because it seems evident that Simon does not control whether he sets off the fire alarm in the same way that he controls whether he flips the switch. Indeed, it's not clear he controls the former at all. This thought can be supported in two ways. First, free will is, after all, about the freedom *of the will* and it is therefore strange to say Simon freely set off the fire alarm when he had no knowledge or awareness that he was doing so. Given that he had no idea how to do it, he could not possibly have willed it. Second, we can point out that there are many other things that Simon (unknowingly) does when flipping the switch, many of which it is even more unintuitive to say that Simon controlled. For example, suppose that when Simon flips the switch, he causes the room's air conditioning unit to become active, raises the temperature of the switch by 0.05 degrees, disturbs some air molecules, and sends a spider scurrying behind a cupboard. Simon has no awareness that he has done any of these things. Yet, according to Ginet's account, Simon exercises the kind of control relevant to moral responsibility over each of these things. In other words, Simon exercises his free will with respect to each of these things; each is something Simon did freely, on Ginet's account. Each of these things is imputable to Simon as an agent in exactly the same way that his flipping the switch is. This, I want to suggest, is implausible. There is a significant difference in the control Simon exerts over his flipping the switch as compared with the control exerted – if it is any form of control – over his raising the temperature of the switch by 0.05 degrees, or sending the spider scurrying.

Proponents of the separation of the freedom and epistemic conditions might be happy to accept these implications. They might suggest that when I ask, doubtfully, whether Simon freely raises the temperature of the light switch by 0.05 degrees, I'm eliciting a negative judgement (if at all) by appealing to intuitions concerning moral responsibility. Behind my denial, they will suggest, is the following reasoning: surely Simon couldn't be *responsible* for raising the light switch by 0.05 degrees, so he cannot have *freely* done this. This reasoning would indeed be problematic because according to the proponents of the distinction moral responsibility requires, in addition to freedom, the satisfying of the epistemic condition. And if the non-responsibility in the above cases is explained by that, then we are not entitled to make any inferences about Simon's freedom or control from his lack of responsibility. As will be seen below, I do not think the judgement I'm eliciting relies on intuitions concerning responsibility. I do, however, accept that the thoughts laid out above do not amount to a knock-down argument against the distinction as drawn. My aim at this stage is just to establish that proponents of the distinction are committed to the idea that Simon was in control of setting off the fire alarm in the same sense that he controlled the flipping of the switch.

I turn now to the task of showing how even those who endorse the distinction place at least some epistemic conditions on the freedom or control condition on moral responsibility. In brief, my argument involves two claims:

(1) Event causation is not sufficient for control or freedom
(2) Whatever else is needed will involve an epistemic condition

I take it that the first point is uncontroversial: human agents are causing all sorts of things all the time, even when they are not acting. Johnny's snoring causes his housemates to wake up. Even if we describe the situation by saying that this was something Johnny *did*, that does not entail that Johnny *acted*. There are lots of things agents do which are not actions: they fall, trip, faint, metabolise, digest food, and so on. So for an event to be an action (if indeed actions are types of events), something more is needed. The second claim is that this 'more' will include an epistemic component. Although I cannot show this to be the case conclusively – to do so would require demonstrating it for each and every possible account of the freedom condition – I think it can be made highly plausible. I take my lead from Mele (2010) who has recently shown how this point holds for Fischer and Ravizza's account of the freedom condition on responsibility. Fischer and Ravizza, recall, endorse the separation of freedom and epistemic conditions, explicitly billing their account as an account of the *control* (i.e. freedom) required – and not of any knowledge required – for moral responsibility (Fischer and Ravizza 1998: 13).

Fischer and Ravizza gloss acting freely in terms of *acting in the absence of undue force* (Fischer and Ravizza 1998: 13) (also see Mele (2010: 103–4)). This gloss confirms the judgement made above that according to their account we should answer in the affirmative to the question whether Simon freely set off the fire alarm and to the question whether Herb freely locked out the plumber. After all, neither Simon nor Herb were subject to any undue force. Simon 'freely' set off the fire alarm inasmuch as he was not being compelled to do so. Herb 'freely' locked out the plumber inasmuch as no one forced him to lock the door. For the same reasons, this initial gloss will require us to say that Simon freely raises the temperature of the light switch and freely sends the spider scurrying.

But as Mele has pointed out, Fischer and Ravizza's detailed account of what it is to act freely, namely, that it is to exercise a certain kind of control, which Fischer and Ravizza label *guidance control*, is in tension with their initial gloss (Mele 2010: 104). This is because, according to Fischer and Ravizza, 'when an agent has guidance control, we assume that he performs the relevant action *intentionally* (i.e. for a reason)' (Fischer and Ravizza 1998: 64). Fischer and Ravizza hold, then, that guidance control is only exercised when the agent acts intentionally and they take the latter to involve a suitable connection between the agent's reasons and the agent's subsequent behaviour: 'the agent's action ... must be intentional, that is, appropriately connect to his reasons' (Fischer and Ravizza 1998: 81).

Mele observes this does not commit Fischer and Ravizza to saying that an agent exercises guidance control over A only if the agent intentionally *A*-s. For the agent might exercise guidance control over A by intentionally *B*-ing. Nevertheless, there is *something* that has to be done intentionally if the agent is to exercise the requisite kind of control. But acting intentionally will involve the agent having significant psychological attitudes. Knowledge might be sufficient, although it is unlikely to be necessary; perhaps some kind of belief is necessary, but that too might be doubted (see Mele and Moser (1994: 44)). The current point, however, is merely that these conditions on intentional action that Fischer and Ravizza invoke to explain when it is an agent acts freely are going to be very similar to the kinds of conditions typically discussed under the banner of 'the epistemic condition on moral responsibility'.

The same point is true of Ginet's account. Agents satisfy Ginet's **CDO** condition if they could have acted in way W and could have not acted in way W. And although his **CDO** condition doesn't explicitly reflect this, Ginet is clear that when he talks about the person *acting in way W*, he's talking about the person *intentionally and voluntarily moving in a certain way* such that various results occur (Ginet 2000: 267). According to Ginet, then, only agents who act intentionally and voluntarily satisfy his **CDO** condition: only agents who act intentionally act freely. And as just emphasised, to act intentionally requires possessing various psychological states and attitudes. Like Fischer and Ravizza's account, Ginet's account allows that agents can freely A without being aware that they are *A*-ing because, again, it allows that an agent freely A-s by intentionally *B*-ing. In other words, according to Ginet, someone might freely A because their A-ing is a result of their *intentionally* B-ing.

According to two prominent accounts offered by those who endorse the separation of freedom and epistemic conditions, then, agents only act freely if they satisfy significant epistemic conditions (those on acting intentionally). And this means that even those who (apparently) endorse the separation of freedom and epistemic conditions in fact require significant epistemic conditions to be satisfied for an agent to exert control. This, I think, raises the following question: would any of those conditions typically thought of as epistemic conditions *on moral responsibility* be better seen as conditions on free will? In the following sections I will argue that at least some of the 'epistemic conditions on moral responsibility' are better thought of as conditions on what an agent is able to control.

5.3. 'Epistemic' abilities

5.3.1. Varieties of 'epistemic' abilities

The stage is now set to ask the following questions: what kinds of epistemic ability are there, and which ones are relevant to free will? Up until now I've been using the phrase 'epistemic ability' to refer to those abilities which require that the agent satisfy some of the conditions typically discussed under the banner of 'the epistemic condition on moral responsibility'. As was noted above, the kind of conditions discussed under that heading are often not *epistemic*: they don't require knowledge but some other psychological attitude. Going forward I will use the term 'epistemic ability' only for abilities which do require the possession of knowledge (typically concerning what the agent is doing). I will articulate accounts of epistemic and doxastic abilities. My primary purpose however is not to develop a full account of all the details of these doxastic or epistemic conditions, but rather to explore in more general terms how fruitful it is to understand these conditions (whatever their exact detail) as conditions that affect the agent's control (rather than as free floating conditions on the agent's being morally responsible). The key feature of the epistemic and doxastic abilities to be discussed is that they require the agent to be able to have a particular representation of the action, namely, a representation whose content (partly) defines the ability.

I will begin, however, by articulating a notion of ability which does not fulfil this requirement. That is, I begin by articulating the kind of ability, which I will call **non-intentional ability**, which those who endorse separate freedom and epistemic conditions on moral responsibility would accept:

(**Non-intentional ability**) An agent S is non-intentionally able to A in circumstances C if and only if
(i) S is able to intentionally B in circumstances C, and
(ii) S's intentionally B-ing in C would be S's A-ing.

The idea with this characterisation of ability is that it captures the kind of ability that is thought to be relevant to free will by those who endorse a rigid separation of the control and epistemic conditions on moral responsibility. In other words, the idea is that the kind of ability Ginet is employing in his **CDO** condition is something like **non-intentional ability**: agents will possess free will with respect to A if they are non-intentionally able to A and non-intentionally able to avoid A-ing. Recall Ginet's example of Simon who, by flipping what looks like an ordinary light switch, unknowingly sets off the fire alarm. According to Ginet, Simon satisfies the control condition with respect to setting off the fire alarm: his setting off the fire alarm is an exercise of his free will. The above account captures this: Simon is non-intentionally able to set off the fire alarm (by intentionally flipping the light switch) and Simon is non-intentionally able to avoid setting off the fire alarm (by refraining from flipping the switch). It's true that to have the non-intentional ability to set off the fire alarm Simon has to be able to do *something* intentionally, but the point is that he does not have to be able to set off the fire alarm intentionally. What marks out **non-intentional ability** as non-intentional, then, is that there is no guarantee that *what it is an agent is said to be able to do* can be done intentionally.

A few further comments on the structure of the above account are in order before proceeding to articulate those notions of ability which require that the agent possess some kind of representation of the action being performed.

First, the account includes the insight from previous chapters that abilities are defined, in part, by a set of circumstances. It is not the ability to A *simpliciter* that is possessed but the ability to A-in-C, where the hyphens here indicate that C is part of the definition of the ability property (I've dropped the hyphens wherever possible to aid readability). In addition, I assume that all the abilities relevant to free will are *maximally specific* (see chapter 3). The 'C' in **non-intentional ability** and in each of the accounts below should thus be understood as being a maximally specific set of definitional circumstances.

Second, **non-intentional ability** is committed to a coarse-grained theory of action individuation such as that proposed by Elizabeth Anscombe (1963) or Donald Davidson (2001). On this view, actions are concrete, particular events which are describable in many different ways. Actions will therefore be intentional only under some description: the action per se is neither intentional nor unintentional (nor even non-intentional). To give an example: Simon's flipping of the switch *is* his setting off the fire alarm – both descriptions pick out the same event. The action is intentional under the former description and unintentional under the latter.

It might be possible to generalise the account be replacing clause (ii) with 'S's intentionally B-ing in C would be, **would be part of, or would result in** his A-ing' (or similar). The first addition, 'would be part of', aims to accommodate those views which see actions are concrete particulars but which individuate them more finely than Anscombe's view. Ginet's (1990) componential view is one such account. He would say, for example, that Simon's setting off the fire alarm is a concrete event composed of the event of Simon's flipping the switch (an action) and the event of the fire alarm going off (not an action) (Ginet 1990: 50). Those two events, one action and one not, are parts of a larger whole which is itself an action. The second addition, 'would result in', aims to accommodate theories of action which take actions to be property exemplifications at times and which therefore tend to individuate actions in a more fine-grained manner. Goldman's (1970) view fits this bill. Goldman thinks, for example, that Simon's action of flipping the switch is the exemplification of the property *flipping the switch* by Simon at a certain time. Given that *flipping the switch* is a different property to *setting off the fire alarm*, Goldman would take these to be two distinct actions. Nevertheless, Goldman thinks these actions are related by what he calls a 'generation' relation. Goldman would say that flipping the light switch, in the circumstances Simon is in, 'causally generates' Simon's setting off the fire alarm (Goldman 1970: 23). Goldman recognises five different kinds of generation relation and the hope would be that the phrase 'would result in' could capture each of these.

Having floated the possibility of such a generalisation, however, I now want to put it to one side. The dominant view in the philosophy of action is that one's theory of action individuation makes little difference to wider issues in action theory and free will. Andrei Buckareff (2011) has recently challenged this view, but even if one's view of action individuation doesn't affect wider issues in action theory, employing such a generalisation would make the ensuing discussion needlessly complex. Mentioning these alternative theories of action individuation is useful inasmuch as it clarifies the nature of ability properties I will be talking about: they are abilities to perform actions conceived of in the way individuated by Anscombe. I turn now to articulating these notions of ability.

Compare **non-intentional ability** with the following:

(**Epistemic ability**) An agent S is epistemically able to A in C if and only if
(i) S is able to intentionally B in C, and
(ii) S's intentionally B-ing in C would be S's A-ing, and
(iii) S knows that by B-ing they would be A-ing.

This notion of ability has an additional condition, namely, that the agent know that by B-ing they would be A-ing. Theorists like Ginet, Timpe and Fischer and Ravizza, who endorse a rigid separation between the control and epistemic conditions on moral responsibility, would take the condition in clause (iii) to be part of the epistemic condition on moral responsibility (See, e.g., Ginet (2000: 275); Timpe (2011: 20–1)). They would consider it irrelevant to the control that the agent exerts. What reason is there for treating it as part of the control condition (i.e. as an aspect of free will)? I've already argued above that Ginet's Simon example shows that knowledge affects an agent's control. The **Trapped Tom** example illustrates this even more strongly.

Recall once more than Tom is trapped in a corridor and is unable to escape because he does not know the combination of the keypad. There is something he cannot control – i.e. whether or not he escapes – and that appears to be due to his lack of knowledge. Now imagine that Tom's friend Tim is trapped on the other side of the building, also in a corridor protected by an electronic keypad. Tim is unfortunate to be in the burning building, but not as unfortunate as Tom because he knows the combination for the lock. And precisely because he has this knowledge, Tim *is able* to escape from the building. It is very plausible to say that Tim has more control than Tom in this situation. More precisely, Tim has control over whether he escapes from the building whereas Tom does not have control over whether he escapes.

The sense in which Tim is able to escape and Tom is unable to escape is captured by epistemic ability: Tim has the epistemic ability to escape from the building; Tom does not have the epistemic ability to escape from the building. To affirm this is not to deny that they also have something in common: both Tom and Tim have the non-intentional ability to escape from the building. This latter kind of ability requires only that the agent be able to do *something* intentionally which would (eventually) lead to his escaping – it doesn't require that the agent know that his escaping would be the result. In other words, the situation can be summarised with the following four statements:

> Tom has the non-intentional ability to escape from the building
> Tom ***lacks*** the epistemic ability to escape from the building
> Tim has the non-intentional ability to escape from the building
> Tim has the epistemic ability to escape from the building

This example, I want to suggest, supports my contention that the kind of knowledge described in clause (iii) is relevant to what the agent controls and thus relevant to the agent's free will. This stands in contrast to those who endorse the traditional separation between free will and an epistemic condition on moral responsibility. For such theorists, the kind of knowledge expressed in clause (iii) is treated as part of the epistemic condition on moral responsibility, and as a result they forced to affirm that Tom and Tim have the same level of control over whether they escape from the building.

Epistemic ability is a strong notion of ability inasmuch as it requires the agent to have *knowledge*. Presumably, unless we are sceptics about knowledge, it will be plausible to think that some agents do have epistemic abilities, at least some of the time. But while epistemic abilities would bestow a robust kind of control, there is good reason to think that such abilities are not required in order for an agent to have free will. Help is available here from a number of sources, most obviously from those who treat clause (iii) as if it were part of a separable epistemic condition on moral responsibility. For my contention is not that those who have attempted to articulate the 'epistemic' requirements as a separate condition have failed to articulate the correct details; rather, my claim is that (at least some aspects of) the condition should not be considered as a separate epistemic condition on being responsible but should instead be seen as part of what it is to have the requisite control. Ginet and Timpe have both pointed out that it is often sufficient for responsibility that the agents truly believe on the basis of good evidence that their B-ing would be their A-ing. This suggests the following account:

> (**Doxastic ability**) An agent S is doxastically able to A in C if and only if
> (i) S is able to intentionally B in C, and
> (ii) S's intentionally B-ing in C would be S's A-ing, and
> (iii) S believes that by B-ing they would be A-ing, and
> (iv) S's belief that they would A by B-ing is true and based on good evidence.

It would be an understatement to say that it is very difficult to spell out what the precise content of the agent's belief needs to be, and what 'being based on good evidence' amounts to. Timpe's (2011: 20–1) final proposed condition runs to one whole page of text and Ginet's is not any simpler. One reason for this complexity is that their accounts are retrospective: given an action performed, they state which prior 'epistemic' (in the broad sense which doesn't necessarily require knowledge) conditions need to have been fulfilled in order for the agent to be responsible. And because beliefs relevant to the attaining of some goal are often acquired while the agent is acting, these accounts have to *trace back* to a time when the agent could have acted so as to acquire the relevant beliefs. For example, suppose that Tim doesn't know the lock's combination but he can easily find it out: there is a book containing all the combinations a minute's walk away (and Tim knows this). Now suppose Tim doesn't go and read the combination for his door: he's busy reading away and has a certain nonchalant attitude to burning buildings, even when he's in them. By the time the flames are licking at his feet, it's too late: his passage to the combination book is now blocked. Intuitively, Tim is at least partly responsible for failing to escape the burning building. And if that's right, then Ginet and Timpe cannot allow Tim's lack of knowledge – at the time at which he is no longer able to escape – concerning the combination to excuse Tim from whatever responsibility he deserves for this. To handle such cases, their retrospective accounts thus need to require either that the agent *has* the relevant beliefs or *could have acted* at some prior time in such a way as to get them. However, when we are articulating forward looking accounts of what the agent is epistemically or doxastically able to do, and not retrospective conditions on the epistemic conditions concerning responsibility, the need for this tracing clause drops out, or rather, is taken care of in the requirement that the agent be aware of some means they have available for attaining their end.

Nevertheless, it is still difficult to fill out the detail in clauses (iii) and (iv). If, for example, what matters is that the agent *intend* to bring about some result, then arguments to the effect that an agent might A intentionally without believing that they are A-ing – Alfred Mele and Paul Moser (1994: 44) present one such argument – would suggest that the belief requirement in clauses (iii) and (iv) is also too strong. However, as my purpose here is not so much as to offer an improvement of the conditions proposed by Ginet or Timpe, but rather to investigate the effect of treating this condition – whatever form it takes – as part of the account of control, I will leave clauses (iii) and (iv) as they are specified in **doxastic ability**, with the understanding that they will need refining in some manner or other.

One point of connection worth making here is with the latest discussion on alternative possibilities in the literature on Frankfurt-style cases. Ostensibly, these discussions concern what it takes for an alternative possibility that an agent has available to be *robust* – that is, what it takes for an alternative possibility to be relevant per se to the agent's moral responsibility. Although cast in this form, these discussions are in effect discussions about the nature of those abilities relevant to free will. And many recent contributions to this literature – especially those by Carlos Moya (2006), Pereboom (2012) and Seth Shabo (2014) – have emphasised not just that robustness has an 'epistemic' component but have provided accounts of this epistemic component in terms of the agent having some awareness of the means they have available to the end in question. These accounts of what it takes for an alternative possibility to be robust thus bear a striking resemblance to clauses (iii) and (iv) as articulated above.

5.3.2. Weak and reliable abilities

The definitions of epistemic and doxastic ability above require that the agent's B-ing *would be* the agent's A-ing. This is a very strong requirement and as such it might be thought rarely satisfied. One modification that might be made is to replace the 'would be' with 'would or would likely be'. This gives the following, which I will call **reliable ability**:

> (**Reliable epistemic ability**) An agent S is reliably epistemically able to A in circumstances C if and only if
> (i) S is able to intentionally B in C, and
> (ii) S's intentionally B-ing in C would or would likely be their A-ing, and
> (iii) S knows that by B-ing they would or would likely be A-ing.

The account of doxastic ability above can be modified in a similar way. If we replace the 'would' with a 'might' we arrive at a notion of ability I will call **weak ability**; again, this modification can be made for the notions of both epistemic and doxastic abilities. Here I write out only that for **weak epistemic ability**:

> (**Weak epistemic ability**) An agent S is weakly epistemically able to A in circumstances C if and only if
> (i) S is able to intentionally B in C, and
> (ii) S's intentionally B-ing in C might be their A-ing, and
> (iii) S knows that by B-ing they might be A-ing.

These modifications are in line with the point made in chapter 3 that abilities are defined, in part, by a modal force parameter: a parameter which, intuitively, represents the 'strength' of the ability. Here I simplify by outlining two notions of ability, reliable and weak, rather than allowing the modal force to vary continuously.[11] In this section I want to ask whether free will requires that agent possess reliable doxastic abilities or weak doxastic abilities. A number of writers have argued that free will requires the former. A similar requirement is often placed on what it is to act intentionally (See, for example, Mele and Moser (1994)). Tomis Kapitan's (1996) account of abilities is one of the clearest expressions of the idea that the abilities most relevant to free

[11] Note that because I'm taking for granted one of my previous conclusions, namely, that the abilities relevant to free will are maximally specific abilities, it will only be possible for an agent to possess a weak ability without also possessing the reliable ability if the universe is indeterministic. The notion of reliability captured by reliable ability, in other words, is one that captures an aspect of the control an agent exerts over a temporal process unfolding in an indeterministic universe.

will need to involve a reliability condition. After stating a number of notions of ability which he says are not what is at issue in the free will debate, he gives the following account of what he refers to as *strict ability*:

> (**Strict**) S is strictly able to bring it about that P if and only if there is a course of action K such that
> (i) S is able to do K, and
> (ii) that P would be a reliable consequence of S's doing K (Kapitan 1996: 424).

Although the account doesn't mention it explicitly, Kapitan is clear that P's being a 'reliable consequence' of S's doing K is meant to be grounded in S's having *the skill* to bring it about that P and also that S has some kind of awareness that their K-ing will bring it about that P (Kapitan 1996: 423–4). Kapitan says that 'responsibility for results implies strict ability to bring about or prevent them' (Kapitan 1996: 424). Kapitan offers the following example in support the reliability condition (understood as requiring skill):

> If I shoot an arrow towards a target 50 meters away circumstances might be such that the arrow hits the bull's-eye. Yet, total novice that I am, I cannot repeat this feat in the next 10,000 tries. ... Were I an expert I might be responsible for the arrow's hitting the bull's-eye – think of the Swiss archer, William Tell, whose obligations were a function of his skill (Kapitan 1996: 423).

In this example, Kapitan says that he lacks the strict ability to hit the bull's eye whereas an expert archer such as William Tell would possess that strict ability. The difference between Kapitan and William Tell is a matter of skill. Moreover, because Kapitan is clear that he thinks William Tell would bear responsibility for hitting the bull's eye whereas he would not, it's clear that he thinks skill is essential to the kind of abilities which are relevant to free will (it's interesting to note than in pursing this line Kapitan is one of the few authors who does not make a sharp separation between freedom and epistemic conditions on being morally responsible, although he nowhere comments on why he thinks his approach is more fruitful).

This same point – the need for a reliability condition on the relevant abilities – is also supported by appeals to the importance of the 'up to us' locution. The 'up to us' locution is often taken to be central to free will. It is sometimes said that an agent has the kind of abilities relevant to free will just when those abilities make it true that the action in question is *up to* the agent (Shabo 2014). We also see this idea in some formulations of the Principle of Alternative Possibilities which explicitly connect the alternatives relevant to moral responsibility with the 'up to us' locution. Alvarez's formulation of this principle is an example of this:

> An agent is not morally responsible for φ-ing at t, unless he could have refrained from φ-ing at t, i.e. unless, at t, it was *up to him* whether or not he φ-ed (Alvarez 2009: 64).

Alvarez takes it for granted that possessing the right kind of abilities will entail that it is up to the agent what they do. This is relevant to the present point because it is unintuitive to describe an action as being up to the agent if it's very unlikely that the agent will succeed in performing the action. If the agent's success rate in performing some type of action is 5%, then we might say that it is up to the agent whether they *try* to perform the action, but we are unlikely to say that the performance of the action is up to them.

Whatever prima facie plausibility this reliability criterion has, however, I will argue that it is mistaken. Before proceeding I need to outline two ways in which it is possible for an agent to do something reliably. First, an agent will (usually) be able to do something reliably if they can perform the action well or skilfully. Second, an agent might be able to do something reliably simply because they can keep trying to do it until they succeed. The difference can be seen by returning to Alice the knot-tier introduced in the previous chapter. Alice is a competent knot-tier, although she doesn't currently know how to tie a Butterfly Knot. Suppose that Alice has a friend, Alison, who is also a competent knot-tier and who *does* know how to tie a Butterfly Knot. There is a sense in which Alice *is not* while Alison *is* able to tie a Butterfly Knot. This sense is one which not only *takes into account* the knowledge and skills Alison and Alice currently have but which actually *holds the knowledge and skills each one has fixed*; that is, this sense of 'able' rules out the acquisition of new knowledge and skills (with one caveat to be mentioned shortly). However, there is another sense in which both Alice and Alison are able to tie a Butterfly Knot. This sense can be emphasised by using locutions such as 'brings it about that' or 'sees to it that'. For example, we might say that Alice is able to bring it about that a Butterfly Knot is tied; indeed, we might say that Alice is able to bring it about that she herself ties a Butterfly Knot. The important point for present purposes is that this would be true even if Alice took four or five attempts to tie the knot. Moreover, given Alice's general knot-tying competence, this will be a reliable ability. The reliability here is attaching to the end result – what Alice brings about – rather than any individual step. My contention is that many ordinary uses of 'able' track this sense of ability and this is because many ordinary instances of agency involve repeated attempts at reaching some goal. Actions in this category can be done reliably, but not skilfully. And there is an ordinary sense of 'able' which tracks this notion because there are often few time constraints on acting, such that often what is most salient is whether some result is brought about (at all) and not how (skilfully) it is brought about. Returning to Alice: there is no question she will tie a Butterfly Knot, even if when she first tries after reading the manual, it takes her a few attempts. Moreover, this is a significant fact about Alice. Someone with no knot-tying skills might not be able to tie a Butterfly Knot even with the manual and plenty of time. Similarly, I can bake a cake *reliably* but not *skilfully*: I will make mistakes along the way and may even have to start from scratch once or twice, but I can (reliably) bring it about that I bake a cake. The motorist who can change a tyre might well take three or four attempts at positioning the spare tyre onto the wheel studs, but that doesn't count against the motorist's (reliable) ability to change the tyre. The infrequent wearer of neck ties might take several attempts to tie the tie, but sooner or later the person will succeed, and so is (reliably) able, in this permissive 'bringing it about' sense, to tie a neck tie.

Once we are clear about the difference between these two ways of having a reliable ability, the idea that free will requires abilities that involve skills becomes much less plausible. To see this, suppose that there is a novice archer similar to Kapitan, call him Tomas, who so lacks skill that he hits his target one out of every ten thousand shots. Moreover, unlike Kapitan, no matter how many shots Tomas fires he never improves in skill. Now, although Tomas isn't skilful enough to simply take aim at his arch enemy and kill her, he does have an enormous amount of of time on his hands (plus lots of spare arrows). Suppose that every day Tomas arrives at his enemy's market stall and tries, time after time, to kill her. On the thirty seventh day, after (say) six thousands attempts, Tomas hits his enemy and she dies. Would anyone be tempted to say that Tomas is not responsible for the death of his enemy because he didn't perform the action skilfully? No. Skill is simply not needed for an action to be imputable to an agent in a way which grounds moral responsibility.

Still, it might be suggested that the second sense of reliability does matter when it comes to free will. Tomas is responsible for killing his enemy, it might be suggested, because he was in control of the end result in the sense that he had the reliable ability to bring about that end result (an ability he has in virtue of having plenty of time and lots of spare arrows). Although this is more plausible than the idea that free will requires skills, it too is false. For example, suppose that Tomas only has a ten minute window in which to fire at his arch enemy. On his second try he hits her and she dies. Tomas's action was an exercise of his free will and the result is imputable to him. But given that he only had a ten minute window in which to fire, he did not have the reliable ability to bring about the end result.

Kapitan's assessment of this kind of case is flawed in two ways. First, the action he considers is that of *hitting the bull's eye*, the performance of which is usually used to assess a person's skill level. Given that focus, it is most natural, when we ask about the person's responsibility, to assess whether there is any skill for which the person deserves praise, and not whether the person is morally responsible for the action performed. We need to be clear, however, that questions about agency, free will and moral responsibility concern a different kind of appraisability to that involved when we praise someone for being skilful.

Second, Kapitan does not make reference to the full range of options that he has available. If we treat Kapitan's firing a shot as something that is *already given* (something that is, in a sense, necessitated), then we will be unlikely to think that Kapitan's hitting the bull's eye is something which he's in control of. This is correct, but it is because, *given his firing of the shot*, Kapitan doesn't have the kind of control to ensure that he hits the bull's eye: by firing the shot he's already done everything he can to bring about his hitting the bull's eye. But if we make clear that Kapitan had available to him the option of not firing the arrow at all, and that he made the choice to take aim and fire at the bull's eye, then we will be more inclined to hold him responsible for hitting the bull's eye even if he cannot do it reliably. This is because, prior to making his choice, Kapitan has control over whether there is any possibility that he hits the bull's eye or no possibility that he hits the bull's eye – Kapitan can ensure either of these outcomes. Again, the action type in question threatens to sway our intuitions here; but we can strengthen the point by altering the example such that Kapitan decides to, fires and hits not a target but rather the heart of his arch enemy.

The following example, which derives from Kane (1996: 55), also emphasises that an agent does not need to be able to reliably perform an action in order exercise the kind of control required in order to be morally responsible (i.e. in order to have free will):

> A plant worker places some radioactive material in his boss's office with the intention of killing his boss. Over a short period of time, before being discovered, the radioactive material emits enough radiation into the worker's boss in order to produce a fatal cancer. It was genuinely indeterminate whether the material would emit enough radiation to give the worker's boss cancer.

I take it to be uncontroversial that the plant worker has the following abilities:

(**A12**) The reliable doxastic ability to place the material in his bosses office
(**A13**) The reliable doxastic ability to refrain from placing the material in his bosses office

It will therefore be fairly uncontroversial to affirm that the plant worker had free will with respect to his placing the radioactive materials in his boss's office. But of course what we want to know is whether the plant worker had free will with respect to killing his boss, and it does not follow merely from the plant worker's possessing **A12** and **A13** that he has any epistemic or doxastic abilities to kill his boss (although it does follow that the worker has the non-intentional ability to kill his boss). The psychological component of such abilities precludes us drawing this inference because it's possible that the worker does not believe that by placing radioactive material in his boss's office he is doing something that might harm his boss. Still, the details of the story make it clear that this is not the case: the plant worker places the material in his boss's office in order to kill him. So the worker has some kind of doxastic or epistemic ability to kill his boss. Which ability? The answer, I think, is the following:

(**A14**) The weak doxastic ability to kill his boss

This is because the worker can do something which might result in his boss's death, but which doesn't make his boss's death very likely. The worker also has the following ability:

(**A15**) The reliable doxastic ability to refrain from killing his boss

If we agree that the worker is morally responsible for killing his boss, then the worker exercised free will respect to that action. This requires the worker be able to perform the action and be able to refrain from performing the action and the considerations above suggest that the kind of abilities here are different: it is enough that the worker have the *weak* doxastic ability to kill his boss, but the worker also needs the *reliable* doxastic ability to refrain from killing his boss. This suggests the following account of the minimal abilities required for free will:

> (**Minimal free will**) An agent has free will with respect to A-ing in circumstances C if
> (i) The agent has the weak doxastic ability to A in C, and
> (ii) The agent has the reliable doxastic ability to refrain from A-ing in C.

Note that this is an account of the abilities which are sufficient for free will, but not necessary. This is because an agent could also possess free will by having two *reliable* doxastic abilities (one to A and one to refrain from A-ing), by having one weak and one reliable epistemic ability, or by having two reliable epistemic abilities. But if an agent has the abilities described in **minimal free will** then the agent exerts the requisite kind of control over A-ing in order to count as being able to freely A. Note too that the 'in circumstances C' in the explanandum refers to the circumstances that the agent is in; this account therefore requires, as per the conclusions drawn in chapter 3, that the agent has two maximally specific abilities that concern the agent's current circumstances. Put another way: the agent will possess the particular (and not just the non-particular) version of the (maximally specific) abilities mentioned in clauses (i) and (ii).

Given that on this account the agent only needs the weak doxastic ability to A in order to exercise free will with respect to A-ing, it might be questioned whether we need a reliability condition on the agent's ability to refrain from A-ing. Perhaps the worker's possession of the reliable ability to refrain was a non-essential element of the example such that the worker might still possess free will with respect to killing his boss even if he only had the weak doxastic ability to refrain from doing so. This, however, is doubtful. Note first that although there is an asymmetry between clauses (i) and (ii) it's not clear the asymmetry is 'deep'. That is, although the plant worker only has the weak doxastic ability to kill his boss, this is because he possesses the reliable doxastic ability to do a number of other things, most relevantly, to place radioactive material in his boss's office. Put another way: although the plant worker only has the weak doxastic ability to *kill* his boss, he has the reliable doxastic ability to *try to kill* his boss. So this example does not establish that an agent could possess free will with respect to some action solely in virtue of having weak doxastic abilities to act in ways that further that goal. Moreover, if we were to consider an example involving a decision and nothing more, such that the agent's two options were deciding to A or deciding not to A, then it's not clear how the weak/reliable distinction is to be applied. The distinction most evidently applies to actions where the agent's activity causes some further consequences which then allow us to re-describe that stretch of activity in various ways. The plant worker's activity involves moving the material into his boss's office and then leaving the office. If the radioactive material does then cause his boss to develop cancer, we can re-describe what he did as his killing his boss (partly because it had this consequence, partly because he intended this outcome). The reliability is a matter of how likely it was that the agent's stretch of activity caused what the agent intended it to cause. Decisions are not like this. The plant worker decides to move radioactive material about in an attempt to kill his boss; this decision won't admit of re-description in the same way on the basis of what the plant worker's own subsequent activity causes.

This is a good result because the reliability condition on the agent's ability to refrain from A-ing accommodates the intuition shared by many that an agent can only be held morally responsible for actions which they could have avoided. As Pereboom says, one reason to affirm that free will consists in possessing the ability to do otherwise is because to be able to do otherwise entails that what you freely do is in some sense avoidable – to have free will is to have the kind of control that would enable you to get yourself 'off the hook' (Pereboom 2001: 1). To be able to get yourself off the hook in this way seems to require that you can *ensure* you did not perform the action; this is part of what makes it fair to blame you for those actions you do which are morally wrong. **Minimal free will** explains this while also explaining cases like the plant worker's where an agent bears responsibility for something despite not being able to ensure success.

One possible downside of the above account, however, is that in virtue of the asymmetry in the abilities minimally required for free will an agent may have free will with respect to some action without it being up to the agent whether they perform that action. To explain: to say that something is up to the agent seems to imply that the agent can reliably do that thing. Intuitively, it's not up to the plant worker whether he kills his boss, and that's because he cannot ensure it happens. If the 'up to us' locution implies this kind of reliability, then when an agent has (and only has) the minimal abilities required for free will (according to the above account), that agent will have free will even though the action in question won't be up to the agent. This is a downside of the view, but, I want to suggest, only a minor one. First, agents often have free will with respect to some action by having two reliable abilities; in these cases, the action will be up to the agent and this helps explain the close connection between free will the 'up to us' locution. Second, even when an agent only satisfies the minimal conditions on free will as outlined above, it's still up to the agent whether they refrain from performing the action in question. This is so because the agent has the reliable doxastic ability to refrain from refraining to kill his boss – an ability the agent has in virtue of being reliably able to try to kill his boss. So the connection between free will and the 'up to us' locution is not entirely severed. Third, and as already mentioned above, although it's true that in such cases the action in question is not up to the agent, many of the steps involved in performing the action will be up to the agent, as will the agent's trying to perform the action. These points, I take it, mitigate somewhat this downside of the proposed account.

5.3.3. Objection: agents who are 'obviously' able to A but who lack the doxastic ability to A

The notions of epistemic and doxastic ability (both the weak and reliable varieties) require that the agent have a specific representation of the action to be performed. If the ability in question is the ability to A, they require that the agent has some representation with the content A. Because agents do not know (or believe on good evidence) everything about the nature of their actions and so don't know each of the different ways that their actions might be described, this opens up the possibility that an agent might have the epistemic (or doxastic) ability to B and yet not have the epistemic (doxastic) ability to A even though their B-ing would be their A-ing. Indeed, recognising this possibility is one reason why we are able to make progress with the assessment of examples such as **Trapped Tom**. But it might be objected that these accounts of abilities also have some counter-intuitive results. Consider the following example:

> (**Penny**) Penny is an accomplished English horn player who has all the skill required to play (one part in) Haydn's 22nd Symphony; moreover, she knows the piece and has played it many times. Penny does not know, however, that this symphony is nicknamed 'The Philosopher'.

On any of the above accounts of ability (excepting non-intentional ability), Penny is not able to perform *the Philosopher*. That is, Penny has neither the reliable nor the weak doxastic ability to perform *the Philosopher*, nor does she have the reliable or weak epistemic ability to perform *the Philosopher*. This is because she doesn't know what 'the Philosopher' refers to. Imagine someone asking Penny to play *the Philosopher*; she'd be at a loss of what to do until she was given or acquired more information. This, it might be thought, spells trouble for the above accounts because it's clearly the case that Penny does have these abilities. Even if Penny doesn't know that the symphony has this nickname, surely it is a mistake to conclude anything about her abilities on the basis of this.

I want to insist, however, that Penny does not have any of the abilities mentioned and that this is no problem for the view even if it is (initially) counter-intuitive. At least two points can be made. First, to deny that Penny has the reliable doxastic ability to play *the Philosopher* is not to deny that she has the non-intentional ability to play *the Philosopher*. Penny has the latter ability because she can intentionally perform Haydn's 22nd symphony and this would be her playing *the Philosopher*. Moreover, Penny can non-intentionally, but with a high reliability, bring it about that she performs *the Philosopher*. So while the current view is committed to denying that Penny has the doxastic and epistemic abilities to play *the Philosopher*, it is not committed to denying that there is no sense in which Penny can play *the Philosopher*.

Second, and as just mentioned, Penny does have the reliable doxastic ability (and presumably the reliable epistemic ability, too) to play Haydn's 22nd symphony. The doxastic ability to play *the Philosopher* is not the same ability as the doxastic ability to play Haydn's 22nd symphony even though someone might exercise each of these abilities by doing the very same thing. Third, we can suggest that it sounds unnatural to deny that Penny has the doxastic or epistemic ability to play *the Philosopher* because this is, in effect, the limit case of a lack of knowledge rendering someone epistemically unable to do something. That is, the only thing Penny lacks is knowledge of the relevant meaning of the phrase 'the Philosopher'. And this knowledge that she lacks is very slight in the sense that although she doesn't know which piece 'the Philosopher' refers to, she knows lots of other things about that piece. In addition, the action in question is very complex such that most people are unable to do it because they lack the skill. In such contexts, when someone does have the required skill, our natural inclination is to affirm that they are able to do it. Moreover, in light of Penny's skill, the very slight knowledge which she does lack seems even more inconsequential than it might otherwise seem. Finally, when it comes to actions such as the performance of a symphony, what we are most likely to be interested in are matters of skill and not matters of free will (as was the case with the archer and the bull's eye). As a result, if the question whether someone is able to play a symphony comes up in ordinary language, the question being asked will almost always be about whether the person has the required skill and not whether they are in a position to play the symphony freely. I think these points lessen any initial hesitation we might have about admitting the notions of epistemic and doxastic abilities and the consequences such accounts bring with them.

References

Alvarez, M. (2009) 'Actions, thought-experiments and the 'principle of alternate possibilities'', *Australasian Journal of Philosophy*, 87/1: 61–81.

Anscombe, G. E. M. (1963) *Intention* 2nd edn. Oxford: Blackwell.

Aune, B. (1967) 'Hypotheticals and 'can': another look', *Analysis*, 27/6: 191–195.

Austin, J. L. (1979) 'Ifs and cans', in J. Urmson and G.J. Warnock (eds.) *Philosophical papers.* Oxford: Clarendon Press.

Ayer, A. J. (1977) 'Freedom and Necessity', in R. Abelson, M.-L. Friquegnon and M. Lockwood (eds.) *The Philosophical Imagination.* New York: St. Martin's Press.

Berofsky, B. (2012) *Nature's Challenge to Free Will.* Oxford: OUP.

Bird, A. (1998) 'Dispositions and antidotes', *The Philosophical Quarterly*, 48/191: 227–234.

Broad, C. D. (1952) 'Determinism, Indeterminism, and Libertarianism: An Inaugural Lecture', in *Ethics and the History of Philosophy*, 195–217. London: Routledge & K. Paul.

Buckareff, A. A. (2011) 'Action-Individuation and Doxastic Agency', *Theoria*, 77/4: 312–332.

Campbell, J. K. (1997) 'A compatibilist theory of alternative possibilities', *Philosophical Studies*, 88/3: 319–330.

Chisholm, R. M. (1964) 'J. L. Austin's Philosophical Papers', *Mind*, 73/289: 1–26.

—— (1966) 'Freedom and action', in K. Lehrer (ed.) *Freedom and determinism*, 11–44.

Choi, S. (2008) 'Dispositional Properties and Counterfactual Conditionals', *Mind*, 117/468: 795–841.

Davidson, D. (2001) 'Agency', in *Essays on actions and events*, 43–61. Oxford: Clarendon Press.

Fara, M. (2005) 'Dispositions and habituals', *Noûs*, 39/1: 43–82.

Ferrero, L. (2009) 'Conditional Intentions', *Noûs*, 43/4: 700–741.

Fischer, J. M. (1979) 'Lehrer's new move: 'Can' in theory and practice', *Theoria*, 45/2: 49–62.

—— (1994) *The Metaphysics of Free Will: An Essay on Control.* Oxford: Blackwell Publishers.

Fischer, J. M. and Ravizza, M. (1998) *Responsibility and Control: A Theory of Moral Responsibility.* Cambridge: Cambridge University Press.

Fischer, J. M. (2012) 'Indeterminism and Control. An Approach to the Problem of Luck', in *Deep Control: A Theory of Moral Responsibility*, 85–105. New York: OUP USA.

Fisher, J. C. (2013) 'Dispositions, conditionals and auspicious circumstances', *Philosophical Studies*, 164/2: 443–464.

Frankfurt, H. G. (1969) 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy*, 66/3: 829-39.

—— (1971) 'Freedom of the Will and the Concept of a Person', *The Journal of philosophy*, 68/1: 5–20.

Gallagher, S. and Zahavi, D. (2008) *The phenomenological mind*: *An introduction to philosophy of mind and cognitive science.* London, New York: Routledge.

Ginet, C. (1990) *On action.* Cambridge: Cambridge University Press.

—— (2000) 'The Epistemic Requirements for Moral Responsibility', *Noûs*, 34/s14: 267–277.

Goldman, A. I. (1970) *A theory of human action.* Englewood Cliffs New Jersey: Prentice Hall.

Gundersen, L. (2002) 'In Defence of the Conditional Account of Dispositions', *Synthese*, 130/3: 389–411.

Hardie, W. F. R. (1971) 'Willing and acting', *The Philosophical Quarterly*, 21/84: 193–206.

Holton, R. (2009) *Willing, wanting, waiting.* Oxford: Oxford University Press.

Honoré, A. (1964) 'Can and Can't', *Mind*, 73/292: 463–479.

Horgan, T. E. (1977) 'Lehrer on 'could'-statements', *Philosophical Studies*, 32/4: 403–411.

Hornsby, J. (1980) *Actions.* London: Routledge & K. Paul.

Huoranszki, F. (2011) *Freedom of the Will*: *A Conditional Analysis* 1st edn. New York: Routledge.

Johnston, M. (1992) 'How to Speak of the Colors', *Philosophical Studies*, 68/3: 221–263.

Kane, R. H. (1996) *The Significance of Free Will* New Ed. New York: OUP USA.

Kapitan, T. (1996) 'Modal principles in the metaphysics of free will', *Philosophical Perspectives*, 10: 419–445.

Kenny, A. (1975) *Will, freedom and power.* Oxford: Blackwell.

Kittle, S. (2015) 'Abilities to do otherwise', *Philosophical Studies*, doi: 10.1007/s11098-015-0455-8 (Online First).

Kratzer, A. (1981) 'The notional category of modality', in H. Eikmeyer and R. Rieser (eds.) *Words, Worlds and Contexts*. *New approches in word semantics.* New York: de Gruyter & Co.

Lehrer, K. and Taylor, R. (1965) 'Time, truth and modalities', *Mind*, 74/295: 390–398.

Lehrer, K. (1968) 'Cans without ifs', *Analysis*, 29/1: 29–32.

—— (1976) ''Can' in theory and practice: A possible worlds analysis', *Action theory*, 1976: 242–271.

—— (1990) 'A possible worlds analysis of freedom', in *Metamind*, 43–79. Oxford: Clarendon Press.

Lewis, D. (1997) 'Finkish dispositions', *The Philosophical Quarterly*, 47/187: 143–158.

Malzkorn, W. (2000) 'Realism, Functionalism and the Conditional Analysis of Dispositions', *The Philosophical Quarterly*, 50/201: 452–469.

Manley, D. and Wasserman, R. (2008) 'On linking dispositions and conditionals', *Mind*, 117/465: 59–84.

Martin, C. B. (1994) 'Dispositions and conditionals', *The Philosophical Quarterly*, 44/174: 1–8.

McCann, H. J. (1975) 'Trying, paralysis, and volition', *The Review of Metaphysics*, 1975: 423–442.

McKenna, M. S. (1998) 'Does Strong Compatibilism Survive Frankfurt-Style Counterexamples?', *Philosophical Studies*, 91/3: 259–264.

Mele, A. R. and Moser, P. (1994) 'Intentional action', *Noûs*, 28/1: 39–68.

Mele, A. R. (2003) 'Agents' abilities', *Noûs*, 37/3: 447–470.

—— (2010) 'Moral responsibility for actions: epistemic and freedom conditions', *Philosophical Explorations*, 13/2: 101–111.

Moore, G. E. (1912) *Ethics.* London: Humphrey Milford; OUP.

Morriss, P. (1987) *Power*: *A philosophical analysis.* Manchester: Manchester University Press.

Moya, C. (2006) *Moral Responsibility: The Ways of Scepticism* 1st edn. London: Routledge.

Mumford, S. (1998) *Dispositions.* Oxford: Clarendon Press.

Nowell-Smith, P. H. (1960) 'Ifs and cans', *Theoria*, 26: 85–101.

Oddie, G. and Tichý, P. (1982) 'The logic of ability, freedom and responsibility', *Studia Logica*, 41/2-3: 227–248.

O'Shaughnessy, B. (1973) 'Trying (as the mental" pineal gland")', *The Journal of philosophy*, 1973: 365–386.

Pears, D. F. (1971) 'Ifs and Cans - I', *Canadian Journal of Philosophy*, 1/2: 249–274.

Pereboom, D. (2001) *Living without free will.* Cambridge: Cambridge University Press.

—— (2012) 'Frankfurt examples, derivative responsibility, and the timing objection', *Philosophical Issues*, 22/1: 298–315.

Pink, T. (1996) *The Psychology of Freedom.* Cambridge: Cambridge University Press.

—— (2004) *Free will*: *A very short introduction.* Oxford, New York: Oxford University Press.

Pollock, J. L. (1976) *Subjunctive Reasoning* 1st edn. Dordrecht, Holland: D. Reidel Pub. Co.

Prior, E. (1985) *Dispositions.* New Jersey: Humanities Press.

Ryle, G. (1967) *Concept of Mind.* Watford, Herts: The Mayflower Press.

Sanford, D. H. (1991) 'Coulds, Mights, Ifs and Cans, Revisited', *Noûs*, 25/2: 208–211.

Schueler, G. F. (1995) *Desire: its role in practical reason and the explanation of action.* Cambridge, Mass., London: MIT Press.

Shabo, S. (2014) 'It wasn't up to Jones: unavoidable actions and intensional contexts in Frankfurt examples', *Philosophical Studies*, 169/3: 379–399.

Steward, H. (2012) 'Actions as processes', *Philosophical Perspectives*, 26/1: 373–388.

Timpe, K. (2008) *Free Will*: *Sourcehood and Its Alternatives.* London: Continuum International Publishing Group Ltd.

—— (2011) 'Tracing and the Epistemic Condition on Moral Responsibility', *The Modern Schoolman*, 88/1/2: 5–28.

van Inwagen, P. (1983) *An essay on free will.* Oxford: Clarendon Press.

—— (1989) 'When is the Will Free?', *Philosophical Perspectives*, 3/1989: 399–422.

—— (2008) 'How to think about the Problem of Free Will', *The Journal of ethics*, 12/3-4: 327–341.

Vetter, B. (2010) 'Potentiality and Possibility', DPhil (Oxford, UK, Oxford University).

—— (2015) *Potentiality*. Oxford: Oxford Univ Press.

Vihvelin, K. (2000) 'Libertarian Compatibilism', *Nous*, 34/s14: 139–166.

—— (2004) 'Free will demystified: A dispositional account', *Philosophical Topics*, 32/1: 427–450.

—— (2008) 'Foreknowledge, Frankfurt, and Ability to Do Otherwise: A Reply to Fischer', *Canadian Journal of Philosophy*, 38/3: 343.

—— (2013) *Causes, laws, and free will*: *Why determinism doesn't matter*. New York: Oxford University Press.

Whittle, A. (2010) 'Dispositional Abilities', *Philosophers' Imprint*, 10/12.

Wolf, S. (1990) *Freedom within reason.* New York, Oxford: Oxford University Press.