Mind Out of Action

The Intentionality of Automatic Actions

Ezio Di Nucci

Doctor of Philosophy The University of Edinburgh

2007

30.9.2007

I, Ezio Di Nucci, hereby declare that I have composed this thesis, that this is my own work, and that this work has not been submitted for any other degree or professional qualification.

Ezio Di Nucci

Abstract

We think less than we think. My thesis moves from this suspicion to show that standard accounts of intentional action can't explain the whole of agency. Causalist accounts such as Davidson's and Bratman's, according to which an action can be intentional only if it is caused by a particular mental state of the agent, don't work for every kind of action. So-called *automatic actions*, effortless performances over which the agent doesn't deliberate, and to which she doesn't need to pay attention, constitute exceptions to the causalist framework, or so I argue in this thesis.

Not all actions are the result of a mental struggle, painful hesitation, or the weighting of evidence. Through practice, many performances become *second nature*. Think of familiar cases such as one's morning routines and habits: turning on the radio, brushing your teeth. Think of the highly skilled performances involved in sport and music: Jarrett's improvised piano playing, the footballer's touch. Think of agents' spontaneous reactions to their environment: ducking a blow, smiling. Psychological research has long acknowledged the distinctiveness and importance of automatic actions, while philosophy has so far explained them together with the rest of agency.

Intuition tells us that automatic actions are intentional actions of ours all the same (I have run a survey which shows that this intuition is widely shared): not only our own autonomous deeds for which we are held responsible, but also necessary components in the execution and satisfaction of our general plans and goals. But do standard causal accounts deliver on the intentionality of automatic actions? I think not.

Because, in automatic cases, standard appeals to intentions, beliefs, desires, and psychological states in general ring hollow. We just act: we don't think, either consciously or unconsciously.

On the reductive side, Davidson's view can't but appeal to, at best, unconscious psychological states, the presence and causal role of which is, I argue, inferred from the needs of a theory, rather than from evidence in the world. On the non-reductive side, Bratman agrees, with his refutation of the *Simple View*, that we can't just attach an intention to every action that we want to explain. But Bratman’s own *Single Phenomenon View*, appealing to the mysterious notion of 'motivational potential', merely acknowledges the need for refinement without actually providing one.

So I propose my own account of intentional action, the 'guidance view', according to which automatic actions are intentional: differently from Davidson and Bratman, who only offer necessary conditions in order to avoid the problem of causal deviance, I offer a full-blown account: *E's φ-ing is intentional if and only If φ-ing is under E's guidance*. This account resembles one developed by Frankfurt, with the crucial difference that Frankfurt – taking 'acting with an intention' and 'acting intentionally' to be synonymous – thinks that guidance is sufficient only for some movement being an action, but not for some movement being an *intentional* action. I argue that, on the other hand, Frankfurt's concept of guidance can be developed so that it is sufficient for intentional action too.

In Chapter One I present and defend my definition of ‘automatic action’. In Chapter Two I show that such understanding of automatic actions finds confirmation in empirical psychology. In Chapter Three I show that Davidson's reductive account of intentional action does not work for automatic actions. In Chapter Four I show that the two most influential non-reductive accounts of intentional action, the Simple View and Bratman's Single Phenomenon View, don't work either. And in Chapter Five I put forward and defend my positive thesis, the 'guidance view'. Also, in the Appendix I present the findings of my survey on the intentionality of automatic actions.

Table of Contents

Acknowledgements

I could not have written this dissertation without the tireless help of my supervisor, Bill Pollard. I also want to thank the many friends, philosophers, students, and colleagues with whom I have talked those things over through the years, in particular Conor McHugh. Finally, I want to thank the University of Salerno for financial support at a vital stage of this process; and my parents, Enza Trotta and Paolo Di Nucci, for all sorts of help, encouragement, and inspiration. I am also indebted to Esmee van Breugel for proofreading this manuscript.

Introduction

… und schreib getrost
"Im Anfang war die Tat"
(Wittgenstein, *On Certainty*, §402)[1]

We think less than we think. Or, anyhow, we think less than philosophers of action think we think. This thesis shows that some prevailing philosophical explanations of human action wrongly appeal to the mind. Those appeals, I argue, are often unjustified and unnecessary. I do so by focusing on what I call *automatic actions*: performances we can effortlessly and successfully complete without paying attention to or becoming aware of them: turning a door handle, *skills* like downshifting to 4th gear, or *habits* like lighting up a cigarette. When we act automatically we don't appear to, consciously or unconsciously, think, nor do we need to think. Through practice, we become confident enough with our automatic performances that we can spare much of the cognitive resources which are normally required by novel or unfamiliar activities.

In Chapter 1 I define automatic actions as performances that we do not nor need to, perceptually or intellectually, attend to; but that are nevertheless within our control. It appears obvious from the examples above that when we act automatically we, normally at least, act intentionally. Firstly, what we do can be attributed to us: it is not just that our body moves; it is rather that we move our body. Also, what happens

[1] Anscombe and von Wright leave the passage in German in the English translation (by Anscombe and Paul) of *On Certainty* (1969), translating it only in a footnote: "… and write with confidence 'In the beginning was the deed'". "Im Anfang war die Tat" is from Goethe, *Faust* I.

is, normally, neither an accident nor a mistake: it is not, therefore, an unintentional action.

When we act automatically, we don't deliberate in advance over whether to act in that way. Nevertheless, we act deliberately. When we act automatically we don't first formulate in our mind a goal that our action is supposed to achieve. Nevertheless, what we do automatically is often goal-directed. When we act automatically we don't reason to decide what we do. Nevertheless, we normally act both rationally and reasonably.

Automatic actions are rational and reasonable all the more because we successfully complete those performances while at the same time resting many of our cognitive faculties (such as, for example, consciousness, attention, control, thought): they are therefore more cost-effective, on average, than the rest of agency. But this is not the only sense in which automatic actions are more effective than non-automatic ones: it is also that we, as agents, are better at our automatic tasks than at our non-automatic tasks. That's part of what it means for a task to become automatic: that we become so good at it that we no longer need to monitor it.[2] The task smoothly runs to completion without bothering our higher faculties (empirical psychology, as I show in Chapter 2, often refers to this as *dual control*).

[2] Just a point of clarification on the fact that we no longer need to monitor automatic performances. It might be argued that in cases such as 'lighting up a cigarette' it is in our best interest to monitor what we do, so that we can hopefully stop ourselves. This would not be a case, then, in which we no longer need to monitor our performance. But here we should distinguish between a sense in which we no longer need to monitor our performance in order to bring it to completion – and this applies to 'lighting up a cigarette' too, since we can successfully light up without monitoring – and a sense in which it is in our best interest to monitor what we are doing. There might be automatic actions which we can successfully execute without having to monitor them, but that nevertheless it was in our best interest to monitor so that we might have been less likely to execute them. 'Lighting up a cigarette' is one such case.

This process of familiarizing oneself with a practice is an essential part of our upbringing. We learn to do things: we learn how to tie our shoe-laces, so that we can soon do so quickly, and without having to first wonder which end goes where. We learn how to button up a shirt, so that we can soon do so without looking at every single button. We learn to swim, and soon enough our four limbs are coordinated, while in the beginning we could only move either legs or arms. This process of learning (which I call *automatization* - a close relative of McDowell's (1994) *Bildung*) can itself be automatic, but need not be.

Acting automatically should be clearly distinguished from other automatic movements of the agent. Our heartbeat might be called automatic, and it is certainly movement: but it is our body that moves rather than us moving our body. It isn't acting. The same goes for other biological functions, but also for unconscious activities like sleepwalking, and for many reflexes. There will be borderline cases, but here I trust that no one will be tempted into arguing that my heartbeat is an action of mine; and that similarly no one will be tempted into arguing that, in normal circumstances, my automatic flipping of a light-switch is *not* an action of mine.

But automatic actions must also be distinguished from all those activities that require care, effort, attention, monitoring: driving for the first time, walking on ice. And from actions that are the result of much hesitation, deliberation, rational weighting of alternative options: signing a big cheque, finally quitting your job. All those actions are not automatic.

I shall leave the task of precisely defining automatic actions, and of distinguishing them from other kinds of movements and actions, to Chapter 1; for the time being, I trust that it is roughly clear to what kind of performances I am referring.

What it is more important to clarify in this Introduction is what this thesis argues for, and what it does not argue for. I shall not argue that automatic actions are intentional actions. I take it to be a crucial intuition of the thesis that when I automatically flipped the switch of the lamp on my desk I intentionally did so.

Given how important this intuition is as a starting point for the thesis I thought it would be worth testing, so I conducted two different surveys. The results of which, presented in the Appendix, show that my intuition that automatic actions are intentional is widely shared. My surveys show that an overwhelming majority of people – above 80% in both surveys - takes humans to act intentionally even when they act automatically. It shows, furthermore, that people don’t tend to distinguish between automatic and non-automatic actions in terms of their intentionality: that is to say, people aren’t more likely to attribute intentionality to a non-automatic action than to an automatic one.

Notwithstanding the survey, I will leave the thesis that automatic actions are intentional as an intuitive assumption, without arguing for it. What I am interested in is, rather, a view of intentional action that can account for the intuition that automatic actions are intentional.

Current accounts, I will argue, fail. They fail because of their commitment to relying on the agent's mind to explain action. When we act automatically our mind is at rest, and it ought not to be artificially 'woken up' at the request of philosophical theories which do not acknowledge the distinctive importance of automatic actions.

The prevailing views of intentional action account for all kinds of actions in terms of causal relations to the agent's mental states. These views divide into two: reductive and non-reductive. Reductive causal views, such as for example Davidson's (1963), take both beliefs and pro attitudes (such as desires) to be necessary for intentional action. While non-reductive views, such as for example Bratman's (1987), appeal to a state of intention which they take to be irreducible to the belief-desire pair put forward by reductionists.

I will show that both the reductive and non-reductive streams of the causalist approach, according to which intentional actions are caused by the mental states (either a belief-desire pair or an intention) that rationalize them, fail to account for the intentionality of automatic action. In Chapter 3 I analyse Davidson's reductive view of intentional action, while in Chapter 4 I analyse two non-reductive views: the *Simple View* and Bratman's *Single Phenomenon View*.

The problem for the philosophical theories that I shall analyse is quickly stated: on these accounts, an action can be intentional only if it is caused by a mental state whose content has a relevant relation to the action in question; as in when I kill JFK with the intention to kill JFK, where my intention to 'kill JKF' causes my action of 'killing JFK'. But when we act automatically there appear to be no such preceding mental states: we just act. And any attempt to superimpose mental states onto the picture will inevitably misrepresent the automatic nature of such behaviours.

Sometimes philosophers have acknowledged this anomaly (in chronological order):

- Whitehead: "It is a profoundly erroneous truism, repeated by all copy-books and by eminent people making speeches, that we should cultivate the habit of thinking of what we are doing. The precise opposite is the case. Civilization advances by extending the number of operations which we can perform without thinking about them. Operations of thought are like cavalry charges in a battle – they are strictly limited in number, they require fresh horses, and must only be made at decisive moments" (1911, quoted in Bargh&Chartrand 1999, p. 464).

- Ryle makes an explicit reference to automaticity: "When we describe someone by doing something by pure or blind habit, we mean that he does it *automatically* and without having to mind what he is doing. He does not exercise care, vigilance, or criticism. After the toddling-age we walk on pavements without minding our steps" (1949, p. 42 – emphasis mine).

- Searle speaks of "actions one performs… quite spontaneously, without forming, consciously or unconsciously, any prior intention to do those things" (1983, p. 84). But I will show in Chapter 4 how Searle's proposed solution is inadequate.

- Bratman too uses the term ‘automatic’: “Suppose you unexpectedly throw a ball to me and I spontaneously reach up and catch it. On the one hand, it may seem that I catch it intentionally; after all, my behaviour is under my control and is not mere reflex behaviour, as when I blink at the oncoming ball. On the other hand, it may seem that, given how *automatic* and unreflective my action is, I may well not have any present-directed intention that I am executing in catching the ball” (1987, p. 126 – emphasis mine). In Chapter 4 I show that Bratman’s *Single Phenomenon View* fails to account for automatic actions.

- Dreyfus says that “expertise does not normally involve thinking at all” (1988, p. 99); and elsewhere he clarifies that what that means is not just the absence of conscious thinking, but also of unconscious thinking: “While infants acquire skills by imitation and trial and error, in our formal instruction we start with rules. The rules, however, seem to give way to more flexible responses as we become skilled. We should therefore be suspicious of the cognitivist assumption that, as we become experts, our rules become *unconscious*. Indeed, our experience suggests that rules are like training wheels. We may need such aids when learning to ride a bicycle, but we must eventually set them aside if we are to become skilled cyclists. To assume that the rules we once consciously followed become unconscious is like assuming that, when we finally learn to ride a bike, the training wheels that were required for us to be able to ride in the first place must have become invisible” (2005, p. 7 – emphasis mine).

- Dennett: “Although we are occasionally conscious of performing elaborate practical reasoning, leading to a conclusion about what, all things considered, we ought to do, followed by a conscious decision to do that very thing, and culminating finally in actually doing it, these are relatively rare experiences. Most of our intentional actions are performed without any such preamble, and a good thing, too, because there wouldn’t be time. The standard trap is to suppose that the relatively rare cases of conscious practical reasoning are a good model for the rest, the cases in which our intentional actions emerge from processes into which we have no access” (1991, p. 252).

- McDowell makes a similar point to Dreyfus’s: “When one follows an ordinary sign-post, one is not acting on an interpretation. This gives an overly cerebral cast to such routine behaviour. Ordinary cases of following a sign-post involve simply acting in the way that comes naturally to one in such circumstances, in consequence of some training that one underwent in one’s upbringing” (1992, p. 50). McDowell’s point can be traced back to a remark of Wittgenstein from *Philosophical Investigations*: “When I obey a rule, I do not choose. I obey the rule *blindly*” (1953, p. §219).[3]

- And even beyond philosophy, here is what Darwin had to say: “It is notorious how powerful it is the force of habit. The most complex and difficult movements can in time be performed without the least effort or consciousness” (1872, p. 35).

[3] Thanks to Paolo for pointing me to this.

What I do in this thesis is best summarized by Dennett's words: I challenge the assumption that a model developed around rare cases of deliberated action – the causal model of Davidson and Bratman - can be applied to all cases, showing that cases such as automatic actions cannot be accommodated by such a model. And then I develop my own account of intentional action, the 'guidance view'; my account does not rely on the agent's conscious or unconscious psychological states causing action.

My 'guidance view' borrows an idea of Harry Frankfurt, *guidance*. Frankfurt, in *The Problem of Action* (1978), criticizes causal views for focusing on the antecedents of action, psychological states, and he proposes to understand actions not in terms of their causal history – whether they have been caused by relevant mental states – but in terms of the relationship between the agent and her actions at the time of action. If this relationship is such that the agent, at the time of action, has guidance over her movements then, argues Frankfurt, those movements are intentional – and they therefore constitute instances of action.

On Frankfurt's view, then, some movement is an action only if it is under the agent's guidance. Frankfurt understands guidance itself in terms of the interventions and corrections that the agent is able to perform over her behaviour, as in the famous car scenario, in which the agent *is driving* her car down the hill even though she is not touching either wheel or pedals just because she is able to *directly* intervene to correct the direction of the car. That ability, according to Frankfurt, is enough for agency.

In presenting my 'guidance view' in Chapter 5 I argue that Frankfurt's guidance is both necessary and sufficient for intentional action. Therefore I go further than Frankfurt, who took guidance to be sufficient only for some movement being an action. On the 'guidance view', then, *E φ-s intentionally iff φ-ing is under E's guidance*.

My view has three basic advantages over causal views such as Davidson's, Bratman's and the Simple View:

- my 'guidance view' can account for the intuition that automatic actions are intentional.[4]
- the 'guidance view' is a full-blown account of intentional action, offering necessary and sufficient conditions, while causal views can only offer necessary conditions.[5]
- on the 'guidance view', as I argue in my Conclusion, the relationship between intentionality and responsibility is much simplified, so that an agent is responsible for all and only her intentional actions.

Why is intentional action important? Why, particularly, is the intentionality of automatic actions important? In short, why have I written this thesis? The concept of intentional action is important: if we weren't able to act intentionally, we probably wouldn't be responsible for our actions. And then, it might be argued, we wouldn't

[4] As I will show in Chapter 5, my view can also account for the intentionality of Hoursthouse's (1991) 'arational actions', which Hursthouse had proposed as a counterexample to the Davidsonian picture.

[5] Therefore my account, differently from causal views, can avoid the problem of deviance.

even be moral agents: therefore it is crucial to have an understanding of what it is to act intentionally. Automatic actions show, I think, that our understanding of intentional action needs refinement. And to such purpose I have written this thesis.

An Introduction is no place for arguments, but there is at least one objection that I must deal with here, because it challenges my motives for working on the intentionality of automatic actions. Automatic actions, it might be argued, are not central to our understanding of agency.

The idea would be that the *important* actions – and the actions that are, for example, more relevant from a moral point of view – are not going to be automatic, exactly because, given their importance and moral relevance, an agent will put much thought, attention, and care into them. Automatic actions, in short, are unimportant – they are the little things of agency: why did you bother?

Indeed, support to this objection appears to come from Velleman's work on action. Velleman (1992, p. 124) has argued that there is an important difference between *half-hearted* actions and *full-blooded* actions: the difference being exactly in the agent's involvement.

> To be sure, a person often performs an action, in some sense, without taking an active part in it… the standard story describes an action from which the distinctively human feature is missing, and that it therefore tells us, not what happens when someone acts, but what happens when someone acts halfheartedly, or unwittingly, or in some equally defective way. What it describes is not a human action *par excellence* (Velleman 1992, p. 124).

Since I have defined automatic actions exactly in terms of the agent's lack of psychological involvement in them, then it looks as though my automatic actions will be cases of Velleman's half-hearted actions; or, anyhow, actions for which it can also be said that they are not cases of action par excellence. The name itself speaks of what Velleman's opinion of them is: they are not central to the agent, they matter little, and philosophy should focus, primarily, on what he calls 'full-blooded' actions. Indeed, Velleman's charge on standard accounts in the philosophy of action is exactly that they can account *only* for half-hearted actions; but that they therefore fail to account for the crucial cases, the full-blooded ones: action par excellence (interestingly enough, if one accepts both Velleman's argument according to which standard causal views cannot account for full-blooded actions, and my argument according to which standard causal views cannot account for automatic actions, then there aren't many actions left for which standard causal views would be able to account).

So here I must reply to the potential objection that automatic actions don't matter much. Support for the claim that automatic actions are important comes from Aristotle's *Nicomachean Ethics* (Book II): there Aristotle's idea is that the virtuous person is the one who *naturally* opts for the good deed; the one who doesn't have to decide or deliberate over which is the good deed. The virtuous deed is, in short, the one that the agent does not need to think about: it is only when virtue becomes second nature that the agent becomes virtuous. The agent, in a slogan, can't *choose* virtue: she must *be* virtuous, as the result of having been habituated to virtue in her upbringing.

> …but if the acts that are in accordance with the virtues have themselves a certain character it does not follow that they are done justly or temperately. The agent also must be in a certain condition when he does them; in the first place he must have knowledge, secondly he must choose the acts, and choose them for their own sakes, and thirdly his action must proceed from a firm and unchangeable character (*Nicomachean Ethics*, Book II, 4 – Ross's translation).

The last sentence is the crucial one: the agent might perform an act that is in accordance with virtue, but if that act does not spring directly out of the agent's "unchangeable character" (her second nature), then her action won't be virtuous. The intuition behind this is, I think, that actions that the agent performs naturally, effortlessly, without hesitation, spontaneously, are *truer* to the agent's self – and many of those actions will probably be automatic ones. Only if the agent's adherence to virtue is true, spontaneous, and genuine can her actions be virtuous and the agent virtuous. Otherwise, according to Aristotle, the agent is merely continent.

This is what it means for automatic actions to be *truer* to the agent's self: they don't tell us who the agent *aspires* to be; they don't tell us what the agent's *ideal* self is (see Smith 1996). They tell us who the agent actually *is*; who she has become through the years; whom she has made herself into.

This is easy to understand: because those actions spring from the agent without the medium of thought, then it is only natural to conclude that they are more *the agent's own* actions than those that have been thought through. The less does an agent think about *φ-ing*, the more is *φ-ing* the agent's own: a *truer* expression of who the agent is, because it is one which wasn't mediated, nor needed to be mediated, by thought.

To understand this it helps to go back to the process of *Bildung* (or automatization): it is a process of internalisation; it is a process of appropriating particular performances that the agent has grown comfortable, and confident, with. Those performances the agent can now make her own: because they represent her particularly well, because she is particularly good at them, or because she particularly likes or enjoys them. This is what it means for something to become second nature: the agent makes it part of who she is.

So the agent develops a particular, special, relationship with some actions rather than others. The idea of *familiarity* comes in handy here: the agent extends her self and personhood to some of her performances but not to others. And what marks those performances as part of the agent's extended self is not that she thinks about them, that she ponders over them, but the very opposite – that she need not think about them.

So automatic actions, half-hearted (or, somewhat more appropriately, 'half-minded') as they might be, are the true heart of one's agency. Here I don't pretend to have developed a conclusive argument against that century-long anti-Aristotelian attitude towards moral behaviour that is often identified with Kantian morality. I just wanted to show that automatic actions are particularly important, especially for someone with Aristotelian leanings.

Here's the structure of the thesis: in Chapter 1 I present automatic actions, distinguishing them from non-automatic actions and from automatic movements that are not actions. In Chapter 2 I show that there is plenty of discussion of automatic actions in empirical psychology; and that what empirical psychologists talk of is precisely automatic actions as I define them in Chapter 1. In Chapters 3 and 4 I argue that standard causal theories of intentional action fail to deliver a satisfactory account of the intentionality of automatic actions: in Chapter 3 I discuss Davidson's view, and in Chapter 4 I discuss the Simple View and Bratman's Single Phenomenon View. Finally, in Chapter 5 I present my own view, the 'guidance view'. Then in the Conclusion I discuss an important consequence of the 'guidance view', that agents are responsible for all and only their intentional actions. The Appendix presents the two surveys I have conducted which show that the intuition that automatic actions are intentional is widely shared.

Chapter 1: Automatic Actions

> It is a profoundly erroneous truism, repeated by all copy-books and by eminent people making speeches, that we should cultivate the habit of thinking of what we are doing. The precise opposite is the case. Civilization advances by extending the number of operations which we can perform without thinking about them. Operations of thought are like cavalry charges in a battle – they are strictly limited in number, they require fresh horses, and must only be made at decisive moments (Withehead 1911, quoted in Bargh&Chartrand 1999, p. 464).

In this chapter I present the subject matter of the thesis: automatic action. I do so in two phases: firstly, I individuate the concept of automaticity; then I distinguish, through the concept of *guidance*, automatic actions from other automatic movements.

1. Automaticity

An awful lot of what we do either is automatic or it involves automatic performances and processes. Think of what you have done so far today: getting out of bed, going to the toilet, putting the kettle on, turning on the radio, brushing your teeth, getting dressed, walking to your office. These, in turn, involved a lot of turning handles, taking steps, raising arms, pushing buttons.

What all those sets of movements have in common is what one could call *mindlessness*: you did not think about these movements, nor did you need to. Mindlessness then distinguishes these movements from others: finally confessing to a wrong-doing; holding on to the rope from which your best friend is hanging; driving through a snow-storm on a mountain road, at night. Those actions are not automatic: they require an awful lot of thinking, wondering, pondering, deliberating, hesitating; an awful lot of attention, care, controlling, making sure. They require both mental and physical effort and strain. Also, those actions, differently from many of our automatic performances, will not be easily forgotten.

But the things that we do automatically described above also appear to differ from other kinds of movements:

- reflexes like eye-blinking;
- tics;
- nervous reactions like sweating;
- biological processes like digestion and heart-beat;
- bodily changes like hair-growth;
- unconscious movements like sleep-walking;
- O'Shaughnessy's (1980, Ch. 10) 'sub-intentional acts', such as the movements of one's tongue.

Even though some of those latter movements seem to be automatic, they don't look like things we do – they don't seem to be actions of ours.[1]

[1] On distinguishing automatic actions from other automatic movements, see also Wright 1976, pp. 127-129.

The aim of this chapter is to establish necessary and sufficient conditions for automaticity, and then to show how automatic actions differ from all these other kinds of automatic movements listed above.

The phenomenon of automatic action has not received much attention from recent philosophy; but when the boundaries between philosophy and psychology were still blurred, there was much talk of automaticity.[2] Here is an excellent example:

> If an act became no easier after being done several times, if the careful direction of consciousness were necessary to its accomplishment on each occasion, it is evident that the whole activity of a lifetime might be confined to one or two deeds – that no progress could take place in development. A man might be occupied all day in dressing and undressing himself; the attitude of his body would absorb all his attention and energy... For while secondarily automatic acts are accomplished with comparatively little weariness – in this regard approaching the organic movements, or the original reflex movements – the conscious effort of the will soon produces exhaustion... It is impossible for an individual to realize how much he owes to its automatic agency until disease has impaired its functions (Maudsley 1873, quoted in James 1890, pp. 113-114).

Maudsley here seems to refer to just the same distinction I drew between automatic and non-automatic actions, when he talks of well-learned practices that have become

[2] My distinction between automatic actions and non-automatic actions might also remind the reader of Collins and Kusch's (1998) sociological theory of action, where they distinguish between *polimorphic actions* and *mimeomorphic actions*. Admittedly, mimeomorphic actions – when agents "intentionally act like machines" (p. 1) – bear some similarities with automatic actions. But my distinction has otherwise nothing to do with Collins and Kusch's.

automatic and tasks that still require the exhaustive contribution of “the conscious effort of the will”. His “secondarily automatic acts” would, then, be what I have been calling automatic actions. Furthermore, Maudsley also accepts that there is a difference between “automatic agency” on the one hand and “organic” and “original reflex” movements on the other, which seems to be just my distinction between automatic performances that appear to be actions and ones that we wouldn’t intuitively call actions.

Contemporary psychology has retained Maudsley’s interest and terminology. Pashler gives us a “widely agreed upon” definition of automaticity:

> At least two changes are widely agreed upon and constitute the core of the concept of automaticity. The first is that practiced operations no longer impose capacity demands, so they can operate without experiencing interference from, or generating interference with, other ongoing mental activities. The second change is that practiced operations are not subject to voluntary control: if the appropriate inputs are present, processing commences and runs to completion whether or not the individual intends or desires this (Pashler 1998, p. 357).

Pashler identifies two features of automaticity on which the literature agrees: lack of capacity demands and lack of voluntary control. Pashler then also lists other features that are often associated with automaticity:

> In addition to these two core elements, many theorists propose that automatic processes have certain additional properties. One of this is functioning without the accompaniment of conscious awareness. Another is requiring little or no mental effort (Pashler 1998, pp. 357-358).

So there are two more features that “many theorists” agree upon: lack of conscious awareness, and lack of mental effort. I should also say that Pashler often refers to lack of capacity demands as “lack of attention demands” (Pashler 1998, p. 377), which gives us a clue as to what those capacity demands that automaticity no longer imposes are.

Pashler here is not only saying that the empirical literature agrees that some features, such as attention, voluntary control, awareness, and effort, are missing from automaticity. He is also suggesting that those features are *no longer required* when an action becomes automatic. So there are at least three distinct points made here: things like attention are not present in automatic phenomena; they are not required in automatic phenomena (anymore); and they stop being required as the agent, through practice, becomes more comfortable with the task, which suggests that actions *become* automatic (or: actions are automatized). This notion of ‘becoming automatic’ is found in Maudsley too: “If an act *became* no easier after being done several times” (Maudsley 1873, quoted in James 1890, pp. 113 (emphasis mine)).

The features individuated by Pashler find confirmation elsewhere. Here is a very good example:

> To start, examine the term automatic… it refers to the way that certain tasks can be executed without awareness of their performance (as in walking along a short stretch of flat, safe ground). Second, it refers to the way an action may be initiated without deliberate attention or

> awareness (as in beginning to drink from a glass when in conversation) (Norman and Shallice 1986, pp. 1-2).

Here, again, we find lack of attention and awareness as the mark of automaticity. Whenever we find the concept of automaticity applied to other behavioural phenomena, those very features are often mentioned. There are two kinds of behaviours that seem to be often associated with automaticity in the literature: skills - playing an instrument, sport, or learning a craft - and habits: turning on the radio in the morning, driving home from work, but also smoking.[3]

Hubert Dreyfus has done much work on skill (1984 - where he and his brother develop a five stages model for skill acquisition - 1988, and 2005). Dreyfus's remarks suggest that his *skilled behaviors* could be automatic: "expertise does not normally involve thinking at all" (Dreyfus 1988, p. 99). Also:

> While infants acquire skills by imitation and trial and error, in our formal instruction we start with rules. The rules, however, seem to give way to more flexible responses as we become skilled. We should therefore be suspicious of the cognitivist assumption that, as we become experts, our rules become unconscious. Indeed, our experience suggests that rules are like training wheels. We may need such aids when learning to ride a bicycle, but we must eventually set them aside if we are to become skilled cyclists. To assume that the rules we once consciously followed become unconscious is like assuming that, when we finally learn to ride a bike, the training wheels that were required for

[3] Expressions of emotions might also be said to be something we do automatically: crying, smiling, blushing, biting one's nails, or shaking one's leg. Those are things we do without thinking; they aren't deliberate; and also it does not look like we pay attention to these performances. Indeed we say of actors, which pay much attention to their facial expressions, that they don't actually or genuinely express emotion, but that they just pretend (they act). In order to keep the examples as simple and uncontroversial as possible, I will not be employing any cases of expressions of emotions throughout the chapter, and I will keep myself to habitual and skilled behaviour.

> us to be able to ride in the first place must have become invisible (Dreyfus 2005, p. 7).[4]

Acquiring a skill means, then, becoming able to do something without thinking about it (without following internalized instructions or rules). Dreyfus's idea that we set aside the rules once we have become skilled resembles the empirical psychology suggestion "that practiced operations no longer impose capacity demands" (Pashler 1998, p. 357). Because there is no longer the need to think about how to do something, the agent's attentional resources are spared – and can, importantly, be deployed elsewhere.

But Dreyfus makes another very important point here: that once we have accepted the phenomenological difference of those kinds of practices, we shouldn't just assume that what used to be conscious (or what in other practices is conscious) is here just unconscious. That itself is part of what Dreyfus calls our "cognitive assumption" (p. 7): it might be that in automatic cases, rather than doing unconsciously what we used to do consciously, we don't think or follow rules at all, consciously or unconsciously (challenging this "cognitive assumption", importantly, will be part of my argument against Davidson in Chapter 3).

With regards to habits, Pollard (2003) actually lists automaticity as one of three features of habitual action:

[4] On this point, see also McDowell: "When one follows an ordinary sign-post, one is not acting on an interpretation. This gives an overly cerebral cast to such routine behaviour. Ordinary cases of following a sign-post involve simply acting in the way that comes naturally to one in such circumstances, in consequence of some training that one underwent in one's upbringing" (McDowell 1992, p. 50).

> What I propose is that a *habitual action* is a behaviour which has three features. It is (i) *repeated*, that is, the agent has a history of similar behaviours in similar contexts; (ii) *automatic*, that is, it does not involve the agent in deliberation about whether to act; and *(iii) responsible*, that is, something the agent does, rather than something that merely happens to her" (Pollard 2003, p. 415).

Pollard's definition of automaticity points to the absence of deliberation. William James says something similar about habits: "habit diminishes the conscious attention with which our acts are performed" (James 1890, p. 114). And Ryle too associates habits with automaticity: "When we describe someone as doing something by pure or blind habit, we mean that he does it automatically and without having to mind what he is doing" (Ryle 1949, p. 42). Ryle's "without having to mind" seems again to suggest that automaticity implies lack of attention or awareness. So we have so far seen automaticity associated with the absence of attention and awareness (Pashler, James, Ryle, Norman and Shallice), thought (Dreyfus), deliberation (Pollard).

Before I analyse the above proposals in order to decide which concepts suit the purpose of defining automatic behaviour so as to distinguish it from non-automatic behaviour, I want to rule out one kind of concept that could be thought to be the mark of automaticity. One might want to propose lack of intention as the mark of automaticity. This seems, indeed, a pretty intuitive idea (and our everyday language seems to support it too).[5] But whatever interpretation of intention one gives, it would not be wise to set lack of intention as a necessary condition for automaticity. If lack of intention is taken to

[5] The results of my surveys – see Appendix – are not conclusive on this issue.

mean that automatic actions are those that the agent doesn't perform intentionally (the so-called *Simple View*, see Chapter 4), then lack of intention, as a criterion for automaticity, would clash with the original intuition with which the thesis started, that things like turning a door handle or downshifting are intentional despite being automatic.

If, on the other hand, one thought that lack of intention did not imply that an action was not intentional, and proposed lack of intention as a criterion for automaticity, then this definition would settle too early the question discussed in Chapters 3 and 4: whether, on influential theories of action like Davidson's or Bratman's, automatic actions are intentional. Those theories rely on intention ('primary reason' in Davidson's terminology) as a necessary condition for intentional action. If I set lack of intention as the mark of automaticity then, by definition, Davidson and Bratman could not say that automatic actions are intentional. I want to find out whether on Davidson's and Bratman's views automatic actions are intentional; I don't want to establish, by definition, and before looking at their accounts, that they aren't. For the same reasons I will not discuss the possibility of the lack of other psychological states (pro attitudes, beliefs) as the mark of automaticity.

One final point for this section. There are different kinds of actions that might turn out to be automatic actions. There are at least two good candidates in the philosophy of action literature[6]: sudden or spontaneous actions on the one side (Malcolm 1968, Davis 1979,

[6] Also, as anticipated in the Introduction, my distinction between automatic actions and non-automatic actions might have reminded the reader of Velleman's distinction between full-blooded action and half-

Searle 1983, Bratman 1987, Wilson 1989, Hursthouse 1991, Mele&Moser 1994), and subsidiary actions on the other (Searle 1983, Brand 1984, Mele&Moser 1994). For the latter kind, Mele and Moser (1994) even talk of automaticity: “In driving to work, an experienced driver shifts gears, checks his mirrors, and the like, with a kind of automaticity suggesting that he lacks specific intentions for the specific deeds. When so acting, he moves his limbs and eyes in various ways, even more ‘automatically’” (Mele&Moser 1994, pp. 231-232). It is important to distinguish between those different kinds because they might pose different problems to standard causal views of action like Bratman's and Davidson's.

Subsidiary automatic actions would be those that are part of some wider action-sequence. So that, for example, if driving is an action-sequence, that will comprise of many subsidiary actions, one of which may be, say, downshifting. Many such actions will be involved in the execution of our habits.

But those are different from sudden actions (or reactions), such as catching a fast approaching ball; or spontaneous actions, such as caressing – or striking - someone. These needn’t be part of any action-sequence (supposing that I am not playing baseball, for example). It needn’t be something that I could reasonably have been expected to

hearted action: automatic actions, then, would be half-hearted in that, for example, the agent does not even pay attention to them; she doesn't even deliberate before embarking in such deeds – they are not, in short, even worth some thought. Velleman (1992) argues that the standard causal view of action might be able to account for half-hearted actions; but it fails to account for full-blooded actions because it fails to include the agent in its explanation.

anticipate, like downshifting while you are driving. Sudden actions will typically be involved in skilled behaviour (think of reaction times in sport).

Both kinds of actions appear to have some, if not all, the features of automaticity individuated in this section: lack of planning, lack of attention, lack of deliberation, and more generally lack of thought. But there seems to be a difference: while subsidiary actions might be easily made to fit into some wider plan of the agent, spontaneous and sudden actions do not necessarily fit into such framework (as Bratman (1987, pp. 126-127) himself admits; more on this in Chapter 4).

This difference might mean that the subsidiary kind of automatic action is easier to deal with for causal views such as Bratman's and Davidson's than the spontaneous and sudden kind, and that’s why I shall keep those two kinds of automatic actions distinct.

Now I shall analyze the different proposals found in the literature to establish which criteria best individuate automaticity.

2. Deliberation

Let us start with deliberation. It could be supposed that what distinguishes automatic actions from non-automatic ones is that the former lack deliberation. When an agent acts automatically, then, she does not deliberate. I think that, by looking at the examples, we can see that this is true. Many, if not most, of our daily activities are undertaken without prior deliberation. In normal circumstances, I don’t ask myself whether to have breakfast, whether to get dressed, whether to go to my classes. But does that mean that I automatically had my breakfast, got dressed, and went to class?

It looks as though many things could have happened while I had breakfast, got dressed, or went to class, as to prevent those activities from being automatic. Suppose, for example, that I found out that there was no milk left; or that I couldn't find my trousers; or that I met an old friend on my way to school. If any of those things had happened, that would not have changed the fact that I had not deliberated whether to have breakfast, get dressed, or go to school. But, given those interferences which spoiled my daily routines, it is difficult to suppose that I have automatically had breakfast, got dressed, and went to school.

This is because there are two distinct levels at play here: lack of deliberation refers to the level of planning, rather than to the level of acting. There is a common attempted solution for this gap: to distinguish between deliberating *whether* to do something, and deliberating *how* to do something (Pollard (2003), for example, draws this distinction for the case of habitual action). Indeed, it might be said that in the cases of interference supposed above, it will remain true that the agent did not deliberate whether to have breakfast or get dressed; but that, given the interference, she will have had to rethink how to go about having breakfast and getting dressed (for example: "There is no milk, so I can't have porridge; I'll have toast instead"; or "I can't find my trousers; so I'll just wear a skirt today"). So it could be said that the agent had to do some deliberating *how*, even though she did not need to deliberate *whether*. And that would explain why the agent did not act automatically: because she had to deliberate how to go about things.

Therefore, it might be proposed that even though lack of deliberation *whether* to φ is not a good enough criterion for automaticity, lack of deliberation *how* to φ is. And so that an action is automatic only if the agent does not deliberate *whether* to φ nor *how* to φ. But there is an easy objection to this: that 'deliberation how' is just a kind of 'deliberation whether' that refers to a narrower action-description[7]. So, for example, it might be that an agent does not deliberate whether to have breakfast, but she has to deliberate how to do it, given that there is no milk. Now, this just means that she has to deliberate whether to have, say, coffee (given that she always takes milk with it), or whether to have toast (given that she can't have porridge). But now the argument just used against deliberating whether to have breakfast can be used against deliberating whether to have toast. And the regress continues.

A way to reply to this objection is to give a different interpretation of 'deliberating how'; one that cannot be reduced back to 'deliberating whether'. This alternative interpretation is, I think, offered by one of the other candidates for defining automaticity: attention. One may say that deliberation how is just attention to the details of action: my attention will be caught by the absence of milk; my attention will be required in order to find my

[7] At this stage it's important that I clarify my position on the action individuation issue. I accept the minimizers' (in the terminology of Ginet 1990) position, according to which different action-descriptions can belong to the same action (position famously held by Anscombe (1957), Davidson (1971), and Hornsby (1980). There are other two main approaches in the literature: maximizers like Goldman (1970, 1971) have it that each action-description individuates a different action. And middlers such as Ginet (1990), Thalberg (1977), Thomson (1979) talk in terms of parts of actions.

trousers; I’ll suddenly realize that the person in front of me is an old friend. If those things happen, one might say, then your behaviour is no longer automatic. Because automaticity requires lack of attention, then those performances are not automatic, because, for one reason or other, I attend to them. So might lack of attention be what characterizes automaticity? Let us then disregard lack of deliberation and turn to lack of attention.

3. Attention

Here is the way in which attention might distinguish between automatic actions and non-automatic actions: take an intuitive case of automatic task such as downshifting when driving. You don’t need to look at the gear-stick; you don’t need to think which gear you want; you don’t need to pay any attention to the whole process: moving your arm and hand down to the left (to the right in *my* car, actually), grabbing the gear-stick, pulling it down, and then bringing your hand back onto the wheel; those are all things you do without paying attention to them. Contrast this with, say, looking for something. Suppose you are looking for your wallet. You will have to think where you have seen it last; you’ll have to think where you usually leave it. But you will also have to go look around for it: on the desk, under the desk, through your clothes, in the kitchen.

Here I have actually spoken of two things that appear to be quite different: looking and thinking. In the literature those two have been identified as different kinds of attention:

> Attention may be divided into kinds... It is either to (a) Objects of sense (sensorial attention); or to (b) Ideal or represented objects (intellectual attention) (James 1890, p. 416).

Ideal or represented objects are, it seems fair to suppose, the objects of thought. This is confirmed by a contemporary version of James's distinction:

> But thinking is not experiencing. There are objects of thought, but an object of thought is not thereby an experienced object, and is not an object of attention in the sense in question (Peacocke 1995, p. 65).

According to James and Peacocke, then, thought is a kind of attention. This means that, in assessing the possibility that lack of attention is the mark of automaticity, we are also discussing thought, which in Section 1 had been individuated as another possible candidate. So the proposal would then be that downshifting is automatic only if the agent does not attend to it. And that would imply that the agent mustn't look at her performance, nor think about her performance.

Let me rule out a first objection: it might be said that, if intellectual attention (thought) is defined by having as its content "ideal or represented objects", as James says, then intellectual attention cannot be about behaviour, because behaviour is not ideal or represented, but real – as in, actual physical movements. So, the objection would go, intellectual attention cannot be part of my definition of automaticity, because physical movements cannot be the content of intellectual attention, which must be "ideal or represented".

The idea seems to be that the content of perceptual attention is the world itself, as in 'I see that there is a pen on my desk'; while the content of intellectual attention, on the other hand (and this would be the difference between the two kinds of attention), must be ideal or represented objects. So perceptual attention and intellectual attention cannot share the same content: if something, as for example an action, can be the content of perceptual attention then it cannot be the content of intellectual attention. But this is just false: 'typing the letter p' can be both the content of my perceptual attention, as in 'I see that I am typing the letter p', and the content of my intellectual attention, as in 'I typed the letter p, but I should have typed q instead'. And in both cases we are referring to the same act-token. So, even though it might be that the same act-token can be differently represented in perceptual and intellectual attention, it looks like it will be the very same act-token which is both the content of our perceptual attention and intellectual attention. So intellectual attention can indeed be about actions.

4. Awareness

So lack of attention looks like a good candidate. But how does it compare with the last criterion that emerged from the literature, lack of awareness? Could lack of awareness be a better criterion for automaticity? Intuitively, attention seems to be the vehicle of awareness: by paying attention (or just *attending*) one becomes aware. This should not be understood as necessarily active: one’s attention might be caught, and then one becomes aware. By looking out of my window (perceptual attention) I become aware that night has fallen. Alternatively a sound catches my attention and I become aware that the phone is ringing. The same is true of intellectual attention: one might recollect a date, or have an idea, and then we'll say that she is (or has become) aware of the date or idea.

These intuitions match what philosophers seem to think. Here is what Roessler says of the relation between awareness and attention:

> A further ingredient in this common-sense understanding of attention is the idea of a connection between attention and awareness. You might tap your fingers on the table, say, without being aware of doing so; and we associate this lack of awareness with the fact that your attention was engaged with other things. But what should we make of this idea? On what might be called a constitutive reading, paying attention to one's action simply is to be aware, in some sense, of what one is doing. Alternatively, on an explanatory reading, the fact that someone pays attention to her action explains that she knows what she is doing (Roessler 2003, p. 389).

On both readings, it seems, one can't have awareness without attention. On the constitutive reading this is because attention just is awareness; and on the explanatory reading because without attention there would be no explanation for awareness. This is not to say that one must have been deliberately paying attention; but it is to say that the agent's attention must at least have been drawn to the object of awareness. This, as we have already clarified, could have happened because the agent drew her attention to it, or because the agent's attention was caught by it.

So it seems fair to conclude that lack of attention either just is or it implies lack of awareness; and that therefore, when we say that an agent who acts automatically does not pay attention to her performance, what we are also saying is that she is unaware of her performance – and that therefore automatic actions are unaware (*unattended* (constitutive reading) or *because unattended* (explanatory reading)).

Still there seems to be a solid intuition that one can be aware of something without paying attention to it. I think it is fair to say that I have been aware all afternoon that there is a door behind me, even though I hadn't paid any attention to it until now. So ten minutes ago it was true that I was aware of the door even though I wasn't paying attention to it. This could be said to be true of both perceptual attention and intellectual attention: ten minutes ago I was not thinking about the door, nor was I looking at it; still I was aware that there was a door behind me. And this seems to contradict the conclusion of the previous paragraph.

This might depend on the fact that perceptual and intellectual attention do not capture the whole of awareness, because they do not capture 'epistemic awareness'. There is a difference (drawn by Dretske (1969, Ch. 2); see also Davis (1982)) between being aware, perceptually or intellectually, *of* the door behind me, and being aware *that* there is a door behind me. This latter kind of awareness we call epistemic awareness. We can clearly be epistemically aware *that* such and such book has a blue cover without being aware, perceptually or intellectually, *of* the book's blue cover; and therefore without attending, perceptually or intellectually, to the book's blue cover. This might be the intuitive sense in which we can be aware that something is the case without paying attention to it. Below I will show that this difference, in the case of action, still does not amount to there being awareness without attention.

The problem can be easily solved if we accept Roessler's explanatory reading: it was not true, even ten minutes ago, that I had never paid attention to the door – sometimes I must have noticed it; either because I attended to it, or because it caught my attention. And the attention that the door received explains the fact that I am aware of the door even now that I am not paying attention to it. But we don't even need to be so strict: there might be facts that I have never actually attended to, but of which I am aware just because I have sometime attended to another fact which implies it.

One might want to say, for example, that at this moment I am aware of the fact that there are fewer than 100 people in the room; even though, clearly, until now I never really formulated that thought. The reason why I am aware of it is simply something else that I will have noticed or thought about: that I was alone, for example (or that there was no one else, or some such thing). If no content relevant to there being fewer than 100 people in the room had ever come to my attention, then we would be lacking an explanation of how I came to be aware of there being fewer than 100 people in the room.

So my commitment to Roessler's explanatory reading means that I am committed to the claim that one cannot be aware of x if one has not attended to (or if one's attention has not been caught by) x or some other fact which implies x. This commitment, in the case of action, does indeed mean that awareness requires attention, because actions are not like doors: when I act, there is no previous attention that can give me awareness of my present acting, because the act-token is happening now and only now. So, with acting, I can only be aware of my acting, at the time of action, if I am paying attention to it, at that time. So we can maintain, for actions, that awareness of action requires attention. And that therefore, if automatic actions are characterised by lack of attention, then they are implicitly characterized by lack of awareness too – and that this also applies to epistemic awareness.

So lack of attention implies lack of awareness. Does it also imply, the reader will ask, lack of knowledge? If I haven't paid any attention to my turning the door handle, then I am not aware of turning the door handle. Given that this lack of awareness implies lack of epistemic awareness, it would appear natural to think that it also implied lack of knowledge. But it seems implausible that I don't know that I am turning the door handle: if you ask me what I am doing while I automatically turn a door handle, I can easily tell you that I am turning the handle. I might, at first, just describe myself as "opening the door"; and so you might need to point to those other activities (exactly because I was unaware of them); but once you have pointed to the fact that I was (also) turning the handle, I will readily admit to it. "Sure, I was (also) turning the handle".

So it looks as though agents know what they are doing despite being unaware of what they are doing. How can this be? Lucy O'Brien has a possible explanation:

> Let us count absent-minded finger tappings as non-intentional actions of mine. Am I epistemologically disassociated from such actions to a degree that makes the claim that I could be totally self-blind with respect to them look plausible? It is clearly true that I can be tapping my fingers without noticing. However, to the extent that it

> is plausible that there is genuine agency in such cases, by which I mean that I can be said to be controlling the action, I must normally be able to come to know what I am doing (O'Brien 2003, p. 365).

This idea of being "able to come to know what I am doing" might well describe the sense in which, when we act automatically, we know what we are doing despite being unaware of it. Should O'Brien's ability to come to know count as genuine knowledge? It doesn't look like reflective knowledge, and this fits both our intuitions about automaticity and the fact that agents are unaware of their automatisms. But I suspect it should still count as knowledge, because to claim that the agent, while automatically turning the door handle, does not know that she is turning it, appears to be plain false.

But I shall leave this epistemological issue to the epistemologist, and so I will not commit myself to the further claim that lack of attention and awareness implies lack of knowledge, even though it probably implies at least lack of reflective knowledge: that is, even if I can be said to know that I am turning the door handle, if I am doing it automatically I am not reflecting upon the fact that I am turning the door handle.

In conclusion, lack of awareness is not a better criterion than lack of attention simply because the latter criterion actually implies the former.

5. Proprioception

One might object to my claim that lack of attention implies lack of awareness on the grounds that there is one kind of awareness that is not implied by attention:

proprioceptive awareness. And that, therefore, I have not shown that lack of attention implies lack of awareness because I have not shown that all kinds of awareness will be lacking: an agent might be proprioceptively aware of what she is doing, the objection goes, even though she is not paying attention to what she is doing.

Proprioception is the subject's awareness of her own body from the inside. The idea is that one is aware of one's own hand independently of one's five senses: "one mode of sense-perceptual access is reserved for the agent's alone, namely the proprioceptive mode" (O'Shaughnessy 2003, p. 348). While I can see my hand, just like someone else can see it, the way in which I can be proprioceptively aware that 'I have a hand' is not available to anyone else. Proprioception depends, as Marcel puts it, on "receptors sensitive to both the interior and the periphery of the body, as opposed to… receptors sensitive to distal stimulation" (Marcel 2003, p. 52). Supposedly, then, when my awareness of my hand depends on the former kind of receptors, we have proprioceptive awareness. When it depends on the latter kind, we have, for example, visual awareness.

So even though it is a form of perception, proprioception appears to be independent from perceptual (as of the five senses) and intellectual attention; and therefore, proposing that automatic actions lack (perceptual and intellectual) attention might imply that they lack perceptual and intellectual awareness, but it does not also imply that they lack proprioceptive awareness. I accept this point, and the clarification it calls for: that when I propose lack of attention as the mark for automaticity, and I say that lack of attention implies lack of awareness, what I mean is that it implies lack of intellectual and perceptual awareness; I don’t mean to claim that, from lack of attention, it also follows lack of proprioceptive awareness. Now, the reader will want to know what the relationship between automatisms and proprioceptive awareness is: does an agent who acts automatically have proprioceptive awareness of what she is doing?

The idea of being proprioceptively aware of what one is doing must be clarified: agents might be proprioceptively aware of the presence and position of their body (see O’Shaughnessy 2003, p. 348), but they can hardly be proprioceptively aware of what they are doing, of their actions. For that they need the aid of perceptual and intellectual attention. Proprioception, as we have emphasized, is of the body and of the body only.

Take our intuitive example of automatic action: turning a door handle. It looks as though an agent can be proprioceptively aware of the forward movement of her hand; she can be proprioceptively aware of her hand being level with her stomach; and she can be proprioceptively aware of the contraction of her fingers. But it is only with the aid of perception (sight, touch) that she can be aware of her hand touching the door handle, turning the door handle, and so on. The action description ‘E turns the door handle’ involves E’s body, but it also involves the door handle. And proprioception can only be about E’s body. Awareness of the door handle will depend on perceptual attention. So we cannot actually say that agents can be proprioceptively aware of their actions.

Here one might want to object that, in ruling out proprioception, I am assuming a distinction that I have not yet drawn: the distinction between actions and other movements. I disagree: firstly, I have not spoken of the agent's movements as opposed to her actions, but of the agent's body as opposed to her actions. Secondly, what I have been employing throughout is just an intuitive idea according to which turning the door handle is both automatic and an action, not the distinction between automatic actions and other automatic movements that I draw in the last section of this chapter. But even if it were true that I had been helping myself to that distinction too early, the distinction itself is not doing any work: the problem with proprioception is that it cannot be said that an agent is proprioceptively aware of 'turning a door handle', but that does not depend on 'turning a door handle' being an action rather than a mere movement; it depends on 'turning a door handle' involving a door handle. Also, for this very reason, proprioception cannot distinguish between automatic actions and non-automatic actions: because in neither case can one say that the agent is proprioceptively aware of her actions. So we can safely rule out proprioception as a criterion for automaticity.

One clarification: here I am not making the controversial claim that actions are constituted by something more than just the agent's movements (Davidson (1971), for example, would deny that). I am just saying that, in most cases, an agent, in order to become aware, or come to know, what she is doing, cannot rely solely on proprioception. Stretching your arm might be a case in which proprioception is sufficient, because it does not involve anything else than just one's own arm.[8] But every time that agents interact with the environment, as with turning a door handle,

[8] Even stretching my arm might not do: it supposedly involves gravitational forces, our awareness of which will depend on our senses.

proprioception isn’t enough to become aware that you are turning a door handle. And it isn’t necessary either: looking at yourself doing it will be sufficient.

Summing up, this section has answered three questions: firstly, I have responded to the possible objection that lack of attention does not imply lack of awareness because it does not rule out proprioceptive awareness. I have clarified that what I take to be implied by lack of attention are lack of intellectual, perceptual, and epistemic awareness, not lack of proprioceptive awareness. And I have shown that one cannot be said to be proprioceptively aware of one’s actions, nor, therefore, of one’s automatic actions. An implication of this is that lack of proprioceptive awareness cannot distinguish between automatic and non-automatic actions, given that it applies to both. It therefore would not do as a criterion for automaticity.

6. Is lack of attention necessary for automaticity?

I have found lack of attention (and awareness) to be a very good candidate as the mark of automaticity. The next question must be whether lack of attention can be a necessary condition for automaticity. To answer this question I must see whether setting lack of attention as a necessary condition would exclude any behaviour that is intuitively automatic.

Is there any attention involving performance that we might still want to call automatic? Ryle (1949) offers one such candidate.

> When we describe someone as doing something by pure or blind habit, we mean that he does it automatically and without having to mind what he is doing. He does not exercise care, vigilance, or criticism. After the toddling-age we walk on pavements without minding our steps. But a mountaineer walking over ice-covered rocks in a high wind in the dark does not move his limbs by blind habit; he thinks what he is doing, he is ready for emergencies, he economizes his effort, he makes tests and experiments; in short he walks with some degree of skill and judgement (Ryle 1949, p. 42).

One might want to describe the movements of the skilled mountaineer as automatic, exactly so that one does justice to the skills of the mountaineer. If that was so, then lack of attention could not be our criterion for automaticity, because clearly the mountaineer is paying quite a lot of attention to his movements. But I don't see any reason to concede this, because I don't see any reason why we should want to say that the mountaineer's movements, in this scenario, are automatic.

In fact, it is Ryle himself who seems to propose a good way of understanding this scenario. Ryle here is distinguishing between two kinds of skills: the skill, which by Ryle's own admission involves automaticity ("he does it automatically"), of a normal walker, and the skill of the mountaineer. The very fact that Ryle is contrasting the two scenarios is evidence for the idea that the latter scenario, differently from the former, does not involve automaticity; but that, Ryle is proposing, does not mean that there is no skill involved in the latter scenario. Sticking to Ryle, then, is enough to dismiss the scenario as a potential counterexample to our proposed necessary condition for automaticity.

There are kinds of habits which might fail to meet the lack of awareness condition, and which might then represent a counterexample to the supposed condition. Think, for example, of virtuous behaviour. Virtuous actions, it has been argued (historically by Aristotle in *Nicomachean Ethics*; for a contemporary defence of the idea, see Pollard 2003), are habitual ones. Think of the pedestrian who unhesitatingly gives to the beggar. If this was indeed an habitual action then, it could be argued, it should be automatic. But it would sound odd to say that the pedestrian was unaware of giving money away.

I don't think this is a counterexample to my proposed condition, because I don't think it is necessary for the action to be automatic in order for it to be virtuous. Pollard (2003) does list automaticity as one of three features of habitual action; but he does not say that it is a necessary feature. And I think it is fair not to set it as a necessary condition, at least for the cases of virtues. I don't think that the agent's noticing that she is giving money away takes away from her virtue (while having thought about it, according to the Aristotelian, does make the agent the less virtuous because of it).

So there don't seem to be any obviously automatic actions left out by setting lack of attention as a necessary condition for automaticity. Lack of attention does so far seem to be shared by all the actions of which we intuitively want to say that we do them automatically. Let us now see whether lack of attention is sufficient for automaticity, or whether we need some further condition.

7. Is lack of attention sufficient?

Would setting lack of attention as a sufficient condition for automaticity include actions that are clearly not automatic? Here it might be proposed that lack of attention is not sufficient for acting automatically because it would include all those other movements listed at the beginning, of which we wanted to say that they don't even count as acting. This might be true: but it must be remembered that here I am only looking for a definition of automaticity, and that it will be for the last section of this chapter to then distinguish automatic actions from other automatic movements. So, for now, all those reflexes and movements should not be used as a counterexample to the sufficiency of lack of attention, because we are interested in its sufficiency for automaticity, and not specifically for automatic action.

A potential counterexample to the sufficiency of lack of attention is represented by unforeseen consequences to what we do, and in general by all those actions that are often called 'unintentional'. It is often the case that an agent, for lack of planning, or lack of attention and care, or just because of stupidity or bad fortune, does something, or brings about something, that she did not mean to do or bring about (or with which she isn't satisfied).

A classic example: suppose I am bitching to Sam about Karl. Suppose that, unbeknownst to me, Karl is in the next room, listening: he gets hurt. Hurting Karl is something I do, and at the same time something I am not aware of – but I am not hurting Karl automatically: in fact, it is not even clear what that would mean, to hurt somebody automatically. Necessarily then, lack of attention (awareness) cannot be sufficient for automaticity, because otherwise my hurting Karl would have to be included, while it is definitely not something I am doing automatically.

It might be proposed, as a way out of this counterexample, that hurting Karl is not an action of mine. That might be, but I can't use that argument here because, again, it is not specifically with automatic *actions* that I am concerned, but only with automaticity. And therefore the fact that hurting Karl might not be an action of mine does not by itself disqualify it from being automatic. I need some other argument to reject this counterexample.

I think that these kinds of cases call for a revision of the lack of attention condition. What characterizes automaticity is not only the absence of attention, but also that automatic performances do not require any attention. With Pashler's words, tasks, when they become automatic, "no longer impose" (Pashler 1998, p. 357) such demands. The agent does not need attention and awareness to complete automatic tasks. This aspect we have already discussed. But it helps us here in distinguishing automaticity from these other things that we are unaware of.

Therefore a second condition must be added: what defines automaticity is not just that there is no attention, but that there is no need for attention. Since both conditions involve attention, it would be tempting to collapse them into one; but this will not do: firstly because, as we have just seen, we can't just say that automatic performances lack attention. But neither can we just say that automatic performances are characterised by the fact that they don't require attention, because sometimes my attention will be caught by what I'm doing, or I will draw my attention to what I am doing, and those performances will therefore cease to be automatic. So we need two conditions: lack of attention and no need for attention. It seems fair to suppose that one condition explains the other, so it will often be the case that agents do not attend to their automatic performances because they don't need to attend to them. Think of downshifting again: you no longer look at the gear-stick because, having grown comfortable and confident with your driving, you no longer need to.

This further condition helps us distinguishing automatic performances from unintentional cases because while in unintentional cases, like the one I presented, lack of attention is the cause of the misunderstanding, in automatic cases lack of attention not only does not compromise the completion of the task, but it in fact promotes the success of the performance. While with mistakes, errors, and unintentional actions it is often the case that they needed more attention – that more attention promotes the success of the performance, with automatic actions it is the opposite: they don't need attention, and often more attention actually disrupts the automatic performance.

Here it could be objected that the success of 'hurting Karl' is also promoted by lack of attention. Had I paid more attention, I would not have hurt Karl – and therefore more attention would have disrupted my 'hurting Karl', just as it can sometime disrupt automatic performances. What is the difference between automatic performances and unintentional performances in terms of attention, then?

It could be proposed that the difference is in the role that attention has played in the history of the performance. Take ‘turning a door handle’. When I was little, I had to learn how to turn a door handle. In the beginning, ‘turning a door handle’ was probably not an automatic action. I had to look at the handle, think whether to turn it left or right, maybe even concentrate. The performance, in short, needed much of my attention. Now I don’t even think which way to turn; and, indeed, if I were to think about it, it would probably take me longer to complete the task, and I might even fail to complete the task more often if every time I wondered which way to turn the handle. This is the history of attention to tasks that have become automatic: it used to be necessary, then it became superfluous, and now it is even counterproductive.

This is not the case with unintentional actions: there was never a time when I learned to do things unintentionally. And there was never a time when I paid attention to my unintentional performances – indeed, those last two statements hardly make any sense. Unintentional performances have always been defined by the fact that I didn’t realize what I was doing (under that description), and by the counterfactual that, had I realized, I wouldn’t have done it – which is not the case with automatic actions. And this is why, now, it makes sense to say that I don’t need to pay attention to my ‘turning the door handle’, but it does not make sense to say that I don’t need to pay attention to my ‘hurting Karl’.

Some psychological literature is quite explicit about the fact that attention is disruptive in automatic cases: "… conscious attention to this aspect of performance can disrupt the action" (Norman and Shallice 1986, p. 3). This is a pretty intuitive idea: try to look at your steps while you walk up the stairs to your office, and you will be more likely to trip over. Similarly, if you are, like me, a compulsive cash machine user, looking at the numbers you are entering often complicates a process you are so good at rather than simplifying it (possibly because you'll inevitably ask yourself, at some level, whether they are the right numbers). But given how many times you have entered that set of numbers, the question is not only unnecessary; it is actually counterproductive: it spoils the automatic flow.

Baseball legend Yogi Berra is reported to have said: "Think? How can you hit and think at the same time?" (Beilock, Wierenga, and Carr (2002), found in Sutton (2007), p. 1). Sutton even refers to this intuition that thought messes up automatic performances as a "prevalent view": "the prevalent view that thinking too much disrupts the practised, embodied skills involved in batting" (Sutton (2007), p. 1).

I think that this idea that attention can disrupt automatic performances could be a good explanation of why the more one practices a performance, the less attention that performance requires. But I don't want to make a further condition of this idea: that attention is no longer required could also just be explained by the fact that agents tend to save unnecessary energies. Also, it is not always the case that, when acting automatically, noticing (or thinking of) what you are doing disrupts your performance. If you happen to look at the gear-stick while downshifting, you are probably not going to get the wrong gear. But while this shows that attention does not always disrupt the performance, it does not show that attention was required; exactly because your attention, we are supposing, is casually caught by what your are doing. If attention were indeed required, then that would mean that you have not practiced enough in order to have made the performance in question automatic.

There is another objection to the sufficiency claim that I must deal with: the reader might wonder why I have excluded Pashler's second condition, which seemed very reasonable: voluntary control. Pashler (1998), as we saw at the beginning of the chapter, claims that it is "widely agreed upon" among psychologists that automatic behaviour is not subject to voluntary control. This seems intuitive: part of some behaviour becoming automatic appears to be that I no longer need to exercise much control over it: I am so used to it that it runs to completion without me needing to check up on it. Think, again, of downshifting: when I downshift from 5th to 4th I don't actually need to voluntarily control that I am downshifting to 4th gear.

Not only the absence of what Pashler calls "voluntary control" is intuitive and supported by the psychological literature, but it is also at the basis of one of the empirical hypotheses that is more favourable to automaticity: the idea of dual control. Dual control, which I discuss at length in Chapter 2, proposes, in short, that there are a conscious level of control and a non-conscious level of control, and that automatic behaviour is characterized by being selected and implemented without the aid of conscious (voluntary) control.

So the reader might object that my two attention conditions cannot be sufficient for automaticity: automaticity also requires lack of voluntary control, as the empirical literature suggests. As I said, I do agree that automatic performances are not subject to the agent's voluntary control, but I don't think that I need to pose lack of voluntary control as a third condition: it just comes with lack of attention and awareness. If an agent were consciously or voluntarily controlling a performance of hers, then it could not be possible for the agent not to be paying attention to or be aware of that performance. That's just what voluntary or conscious control of a movement is: attention to that movement such that the agent is aware of that movement. So when I say that some automatic performance is unaware, it follows that the agent is not consciously or voluntarily controlling it (which, as dual control suggests and as I show in the last section of this chapter, does not mean that the agent is not *in* control).

The same is also true of another condition mentioned by Pashler: effortlessness. It is probably true that automatic actions lack mental effort, but there is no need to make that into a further condition because if an action gave rise to mental effort, then that would presumably catch the attention of the agent, and so the action would not be automatic on grounds, again, that it would not meet the lack of attention condition. So effortlessness, as voluntary control, follows from lack of attention.

There is at least another intuitive feature of automaticity that is implied by lack of attention: the fact that agents can't remember many automatic actions. Presumably, if the agent did not attend to her performance, she has no way of remembering her performance. I don't remember turning the door handle the last time I came into the office; and that's probably because I didn't pay attention to it. Evidently, here I am not proposing that every activity of ours that we can't remember is automatic; nor even that every recent activity of ours that we can't remember is automatic. Only that, if one thought that it was characteristic of automatic actions not to be remembered by the agent (and I am not going to commit myself to this further claim here), then there would be no need to propose this as a further condition for automaticity, because it just follows from lack of attention.

In conclusion, I think that there are two individually necessary and jointly sufficient conditions for automaticity: some behaviour is automatic if and only if the agent does not need to attend to it and the agent does not attend to it. Having now defined automaticity, the job of the last section is to distinguish between automatic actions and other automatic movements.

8. Guidance and Intervention control

The aim of this section is to distinguish two different phenomena that meet the conditions for automaticity: on the one side, there are movements such as down-shifting or turning a door handle, of which we want to say that they are actions; and therefore that they are automatic actions. On the other side there are all those movements mentioned in Section 1, such as eye-blinking, heart-beat, or sleep-walking which, despite being automatic, we don't normally call actions.

There is one possible way of drawing this distinction between automatic actions and other automatic movements that I want to rule out at the outset: I could just help myself to one of the established criteria of action, such as, for example, Davidson's idea that some movement is an action only if it is intentional under at least one description (Davidson 1971). The reason why I won't use such a well-established account of action is that in Chapter 3 I will argue that Davidson's account is not suitable for automatic actions; so I can't help myself to it now.

Harry Frankfurt (1978) has famously argued that agents don't need to voluntarily or consciously control what they are doing in order to be *in* control of what they are doing. Frankfurt calls this idea of 'being in control' without 'controlling' *guidance*:

> A driver whose automobile is coasting downhill in virtue of gravitational forces alone might be satisfied with its speed and direction, and so he might never intervene to adjust its movement in any way. This would not show that the movement of the automobile did not occur under his guidance. What counts is that he was prepared to intervene if necessary, and that he was in a position to do so more or less effectively. Similarly, the causal mechanisms which stand ready to affect the course of a bodily movement may never have occasion to do so; for no negative feedback of the sort that would trigger their compensatory activity might occur. The behaviour is purposive not because it results from causes of a certain kind, but because it would be affected by certain causes if the accomplishment of its course were to be jeopardized (Frankfurt 1978, p. 160).

Some movement can be under the agent's guidance even though the agent is not actually causing the movement in question, nor doing anything in order to control the movement. So an agent can be in control of movements that she is not herself causing: gravitational forces, rather than the driver, are causing the car's movements in the scenario. But nevertheless the car is under the agent's control. This is because at any time the agent can intervene to redirect the car's movements. In applying this concept to habits, Pollard (2003) calls it *intervention control*:

> For we have the capacity to intervene on such behaviours. This is particularly the case for those automatic behaviours which we have learned. Since there was a time when we didn't do such things, it will normally still be possible for us still to refrain from doing them in particular cases (though perhaps not in general). We intervene by doing something else, or nothing at all, either during the behaviour, or by anticipating before we begin it. In this way habitual behaviours contrast with other automatic, repeated behaviours such as reflexes, the digestion, and even some addictions and phobias in which we cannot always intervene, though we may have very good reason to do so. I call this *intervention* control (Pollard 2003, p. 416).

From Pollard's remarks it is clear how helpful the concepts of guidance and intervention control can be for me in distinguishing automatic actions from other automatic movements.[9] Automatic movements such as eye-blinking, heart-beat or sleep-walking, it is easy to see, have two crucial features in common with automatic actions like turning a door handle or downshifting: they are movements of my body; and they are unaware.

[9] In Chapter 5 I highlight some differences between guidance and intervention control – namely, that the agent's capacity for intervention is a pre-requisite for guidance, but that guidance is a more specific concept than intervention control. But for the time being those differences are not crucial.

But guidance gives us a very good way of distinguishing between those automatic movements that we want to call actions, and those that we don't.

The idea is that while with automatic movements such as turning a door handle we can always (and easily) intervene to stop ourselves from performing them, the same cannot be said of automatic movements such as heart-beat and eye-blinking. We don't, I take it, have guidance over those latter movements. This does not mean that it is always impossible, for us, to avoid blinking our eyes. We know that, if we try hard enough, we can avoid blinking for a while. But we can't avoid blinking for good, and we can't always avoid blinking; the same way in which we cannot avoid breathing for good (assuming that killing ourselves does not count as a way of controlling our breathing patterns).

Here we must distinguish between a *direct* way of intervening, and an *indirect* way of intervening. In fact, it is not impossible for me to stop myself from sleep-walking. I can lock the bedroom's door or tie myself to the bed. But this looks very different from stopping myself from turning the door handle. Indeed, I want to say that only the latter kind of control counts as guidance. Tying myself to the bed is an indirect kind of control which I don't think is sufficient for guidance. One way of drawing this intuitive distinction is by saying that while, in order to control my sleep-walking, I need to do something else - tying myself to the bed - there is nothing else I need to do in order to stop myself from turning the door handle.

This difference might not be enough to establish what an action is, nor to establish that automatic behaviours such as turning a door handle do indeed count as actions, but it might very well be enough to show why automatic movements such as eye-blinking, heart-beat and sleep-walking are not actions.[10]

There is, nevertheless, a possible objection to my use of guidance to draw the distinction between automatic actions and other automatic movements: that guidance is not compatible with automaticity because an agent cannot have guidance over a performance of which she is unaware. To see this objection, take Frankfurt's own scenario: for the driver to have the ability to intervene over the car's movements if she wants or needs to, the driver must be aware of the car's movements. If she is unaware of where the car is going, the driver cannot intervene in order to redirect it.

The same point can be made if we come out of the metaphorical scenario and abandon the car: if the agent is not paying attention to her own door handle turning, the objection will go, how can she possibly have the ability to intervene to stop herself from turning the handle? How can she possibly have the ability to stop herself from doing something that she is not aware of doing?

The answer is already suggested in Pollard's passage that I quoted earlier: the fact that agents are acting automatically does not mean that they don't have the ability to draw

[10] In Chapter 5 I actually argue that guidance is sufficient for agency; and that will have the interesting consequence that not all actions of the agent are caused by the agent.

their attention to what they are doing, if they want or need to. The claim is that lack of attention characterizes automatic action. It does not follow, from the claim that automatic actions are not attended to by the agent (and that the agent does not need to attend to them), that the agent does not have the ability to attend to them. Take the case of downshifting again: the fact that you do it without paying attention to it and the fact that attention is not required do not imply that you cannot, if you want or need to, draw your attention to the performance. Acting automatically does not mean that your ability for attention is impaired; only that it is spared.

To show that agents, while acting automatically, have such an ability to draw their attention to something at will, it will help to compare acting automatically with those other automatic movements. Take, for example, sleep-walking. While you are sleep-walking, you might happen to wake up, and wonder what you are doing in the staircase. But you can't wake up at will; that is not the way in which sleep-walking works. So, in the case of sleep-walking, the agent does not have the ability to draw her attention to something at will – for example, the agent does not have the ability to draw her attention to the fact that she is sleep-walking at will.

Contrast this with our standard cases of automatic actions: downshifting and turning a door handle. The fact that you are not normally paying attention to those performances does not mean that you can't, if you want or need, draw your attention to them. And if you can draw your attention to them, as in the case of downshifting and turning a door handle, then you can, normally, stop yourself from doing it. On the other hand, if you cannot draw your attention to your performance, as in the case of sleep-walking, then you cannot stop yourself. This is, then, the sense in which automatic actions like downshifting or turning a door handle are different from automatic movements such as sleep-walking.

We might want to conclude, then, that having the ability to become aware of, or draw one's attention to, some action of ours is one way in which we can be said to have guidance over that action. But it needn't be the only way. Automatic actions being performances with which we have much familiarity and practice, it seems likely that, were an error or anomaly to occur, our attention would be caught by the fact that the usual pattern was being spoiled.

Indeed, it might be argued that part of the process of an action becoming automatic (compare McDowell's *Bildung* (1994) – on this, see also Chapter 3, Section 3.3) is the agent's acquiring the ability to detect anomalies without any active participation on the part of the agent herself. So that the agent does not need to be 'on the lookout' for anomalies; she doesn't need to be paying attention in case anomalies were to occur. If and when anomalies did occur, they will inevitably catch the agent's attention, because, given her familiarity with the pattern, the difference will be too striking to go by unnoticed. Indeed, were an agent not to detect the anomaly, we would say that she hadn't yet learned her craft properly: that the practice had yet to become *second nature*.

So it is the anomaly that does the work by catching the agent's attention; the agent doesn't need to do anything. But, obviously, for something to count as an anomaly, the agent must have done something in the past: namely, she will have habituated herself to a practice in such a way that, if something new happens in the context of that practice, it will count as an anomaly – and catch the agent's attention.

There is another possible objection I must deal with: it could be argued that I have only shown that agents can intervene to stop themselves during the performance; but that I have not shown that agents can stop themselves before the performance. And, the objection could go, if agents do not have this sort of *prior* control over what they do, then they don't have control at all.

The idea is that an agent can draw her attention at will to what she is doing and, if she likes or needs to, stop herself. But how can agents draw their attention to something that they are not yet doing? If you can draw your attention to the fact that you are turning a door handle, you can also draw your attention to the fact that you are approaching the door, or about to turn the door handle. And, I want to say, if in the former case your ability to draw your attention to what you are doing means that you can stop yourself, then in the latter case your ability to draw your attention to what you are about to do means that you can stop yourself from doing it. The latter case, admittedly, needs something else: possibly a belief, judgement or expectation on what you are about to do; so that you can infer, from, say, the fact that you are approaching the door, that you are about to open it.

So, while in the former case the only judgement required might have been the realization that you are turning the door handle, in the latter case the agent needs, on top of the judgement that you are approaching the door, a judgement like “therefore I must be about to turn the door handle”. The presence of these judgements implies, in the former case, perceptual attention and, in the latter case, both perceptual attention and intellectual attention. But the presence of these judgements is no problem for automaticity: when you intervene, either to stop yourself while you are doing something or to stop yourself *from* doing something, the action ceases to be automatic (or is no longer automatic). Importantly, the ability to intervene is only dependent on the capacity to become aware, rather than on previous, or constant, awareness. And the agent need not have those beliefs and expectations while she is acting automatically; she only needs to be able to make these judgements once she has become aware.

I want to deal with one final objection. Flexing one’s muscles, an objector might say, is under the agent’s guidance, but it is not, intuitively, an action of ours. I don’t want to take issue with the idea that we have guidance over flexing our muscles: we can intervene over our muscle flexing, to flex more or less, or to stop flexing. What I don’t see is why it would be problematic to say that flexing our muscles is an action of ours. The objector might propose that we don’t want to refer to flexing as an action because it is never something we just do, but always something we do in order to act. It is a functional prerequisite or part of acting, but it is not itself acting. But if we concede that flexing is something we do, I don’t see what should stop us from conceding that it is an action.

It is definitely different from the mere bodily movements from which I want to distinguish actions through guidance: for a start, it doesn’t happen to us, we do it. Also, flexing our muscles might not be the only thing that we never just do: it might be that we never just move our arm; that there is always some other description of our moving our arm. Further, as Hornsby (1980 – Anscombe too (1957)) points out, we can devise a scenario in which we just flex our muscles: “A man learns that certain particular muscles of his arm have to be contracted if ever he is to clench his fist; and we may imagine that he has a reason to contract those very muscles – perhaps he wants to please some experimenter. He does so. As we say: he contracts his muscles by clenching his fist” (Hornsby 1980, p. 20).

Indeed, this appears to be a case in which it is fair to say that the man, in contracting his muscles, acts. And Hornsby later suggests that, on the Davidsonian view of individuation that I accept (one on which different action-descriptions can belong to the same action, as opposed to Goldman’s view (1970) that every action description individuates a different action – see footnote 7, this chapter), flexing one’s muscles might indeed be an action of ours: “On the view of individuation I have argued for, there is only one action when a man contracts his muscles by clenching his fist, a single performance on his part” (Hornsby 1980, p. 22).

On this reading, then, clenching one's fist and contracting one's muscles are two different action-descriptions of the same action. Davidson would argue, further, that those two descriptions are both descriptions of actions rather than of mere bodily movements because they are intentional under at least one description; which here would actually be 'contracting one's muscles' (while in normal cases it would be 'clenching one's fist'). As I said before, I won't commit myself to this latter part of Davidson's view, but I don't see why, if Davidson accepts that contracting one's muscles can be an intentional action, *I* should deny that it can be an action in the first place.

So, in conclusion, the difference between automatic actions and other automatic movements is that agents have guidance over their automatic actions, but they don't have guidance over other automatic movements such as heartbeat and sleepwalking. This ability to intervene *directly* depends on the fact that agents have the ability to draw their attention to their performances at will.

Conclusion

In this chapter I have presented the subject matter of the thesis: automatic action. I have done so in two steps: first, I have defined the concept of automaticity; and then I have distinguished between automatic actions and other automatic movements. In defining automaticity, I have proposed two individually necessary and jointly sufficient conditions: some behaviour is automatic iff the agent does not need to attend to it and the agent does not attend to it. I found that lack of attention is a better criterion than lack of deliberation, which would include many performances that are intuitively not automatic. I also found that lack of attention implies lack of awareness, effortlessness, and lack of voluntary control. So I did not need to set those as further conditions for automaticity. After having defined automaticity, I have distinguished between two sets of movements which meet the conditions for automaticity: tasks such as turning a door handle or downshifting when you are driving, which we intuitively consider actions; and movements such as heart-beat and sleep-walking, which we don't intuitively refer to as actions. I have shown that the difference between those two kinds is the agent's guidance: agents have the ability to intervene directly on performances such as turning a door handle, but they don't have the ability to intervene directly on movements such as sleep-walking.

In the next chapter I will show that automaticity and automatic actions have received much attention from empirical psychology; and that the kinds of automatic behaviours empirical psychologists are interested in meet my conditions for automatic action as set in this chapter.

Chapter2: Automatic Actions in the Empirical Literature

Introduction

The aim of this chapter is to show that the concept of automatic action that I have defined in Chapter 1, despite being novel within philosophical literature, has been well-established in empirical psychology for decades. I will do this by looking at the work of Bargh on automatic biases and at the work of Norman and Shallice on dual control; and I will show, crucially, that both are talking about automatic actions in my sense.

Why should a philosopher even bother establishing the presence of a philosophical concept within empirical research? First of all in order to show that the concept of automatic action matters, and that its interest and relevance reaches far beyond philosophy.[1]

Also, if one accepts that empirical hypotheses and data can constitute evidence in favour of a conceptual claim, such as my definition of automatic action from Chapter 1, then this chapter will also provide further justification for my definition. But here I won't defend the antecedent of the above conditional.

[1] And, given the lack of philosophical discussion of automatic action, in order to show that I am not the only one who is interested in the topic.

It is important to clarify that the chapter does not mean to provide an argument for the existence of automatic actions: I take what I say in Chapter 1 to be sufficient for that purpose. Nor does the chapter mean to present a causal explanation of automatic action: here it is (only) in the definition of the concept of automatic action that I am interested.

1. Bargh

Bargh's work is probably the most influential research on the topic of automaticity. Bargh (1996, 1999) has conducted notorious experiments on the extent of automatic influences over human behaviour. In this section I will show that the kinds of behaviours which result from Bargh’s experiments meet my conditions for automatic action.

But first I think it is important to provide some background on Bargh’s motivation for studying automaticity. Bargh, like me, takes automaticity to be a *good* phenomenon: he is interested in the way in which automaticity makes us more successful agents, in that it increases our familiarity with a task, and decreases the energies required to complete it:

> Thus “the automaticity of being” is far from the negative and maladaptive caricature drawn by humanistically oriented writers; rather, these processes are in our service and best interests – and in an intimate, knowing way at that. They are, if anything, “mental butlers” who know our tendencies and preferences so well that they anticipate and take care of them for us, without having to be asked (Bargh and Chartrand 1999, p. 476).

Differently from me, though, Bargh appears to think that automaticity poses questions of free will.[2] This is how Bargh himself describes his interests: “My lines of research all focus on the question ‘How much free will do we really have?’” (from his web-page: http://bargh.socialpsychology.org/). In this section, by arguing that the behavioural responses of Bargh’s experiments are automatic actions, I deny that Bargh’s cases pose more of a problem to free will than do any other kinds of action.

Let us look now at the most famous of Bargh’s experiments, in which participants were primed with “words related to the stereotype of the elderly (e.g., Florida, sentimental, wrinkle)” (Bargh and Chartrand 1999, p. 466[3]), while a control group was primed “with words unrelated to the stereotype” (ibid). The findings were quite amazing: “participants primed with the elderly-related material subsequently behaved in line with the stereotype – specifically, they walked more slowly down the hallway after leaving the experiment” (ibid) [4]. Importantly, subjects were not aware that they were primed with the elderly stereotype: “No participant expressed any knowledge of the relevance of the words in the scrambled-sentence task to the elderly stereotype” (Bargh, Chen, and Burrows 1996, p. 237).

[2] Just like Libet (1985) took his experiments’ results to pose a challenge to free will.

[3] The full list of words: “worried, Florida, old, lonely, grey, selfishly, careful, sentimental, wise, stubborn, courteous, bingo, withdraw, forgetful, retired, wrinkle, rigid, traditional, bitter, obedient, conservative, knits, dependent, ancient, helpless, gullible, cautious, and alone” (Bargh, Chen, and Burrows 1996, p. 236).

[4] According to my own calculations on Bargh’s (1996) data (the article does not give the speed), the group subject to the elderly stereotype does on average a speed of, approximately, 4.2 kph, while the control group does 4.8 kph.

The agent's behaving under the influence of the elderly bias appears to meet my conditions for automaticity: the agent does not need to attend to the fact that she is subject to the bias (or, more specifically, to the fact that she is behaving under the influence of the bias), in order to so behave. Also, the agent does not indeed attend to the fact that she is subject to the elderly bias - and that, supposedly, explains why the agent is not aware of the bias. So the agent's behaving under the influence of the bias does indeed meet my two individually necessary and jointly sufficient conditions for automaticity.

This remains true if one changes the description of what the agent is doing: 'walking at 4.2 kph', for example: the agent *can* do that without paying attention to what speed she is doing; and she probably *will* do that without paying attention to what speed she is doing. So this description meets my conditions on automaticity too.

The remaining question is whether what the agent does is an automatic *action*: so whether the agent's movements, in leaving the experiment's room, are under the agent's guidance or intervention control. In order to answer this question, we must first establish what it is that the agent does.

This is because even though, as I have already said in Chapter 1, I accept, with Anscombe (1957), Davidson (1971), and Hornsby (1980), the view that different action-descriptions can belong to the same action, that does not mean that different descriptions of the same action share all the same properties. Take, for example, Davidson's case

(1963) of flipping the switch to turn the light on and, unbeknownst to me, alerting the prowler. According to Davidson, ‘turning the light on’ and ‘alerting the prowler’ are different descriptions of the same action, but that does not mean that they have the same properties: ‘turning the light on’ is, for example, intentional on Davidson’s account; while ‘alerting the prowler’ isn’t.

It might be proposed, here, that the property of being intentional works at a different level from the property of being under the agent’s guidance; because the former is a property of action-descriptions, while the latter is a property of movements; and that a consequence of this is that all action-descriptions that belong to the same action necessarily share the property of being under the agent’s guidance, while they don’t necessarily, as we have seen, share the property of being intentional.

The reason why I won’t help myself to this point here is that I will challenge this point in Chapter 5; so now I can’t let my discussion rely on a claim that I will later reject. But I will show that Bargh’s cases are automatic actions even without the aid of the above claim.

So to establish that one description has some property – being under the agent’s guidance, say – does not necessarily imply that another description of the same action has that property. Therefore, if all I could show was that the description ‘walking at 4.2kph’ is an automatic action, then it might be objected that I have not shown that Bargh’s cases are automatic actions, simply because the description ‘walking at 4.2kph’ does not actually capture Bargh's cases because it leaves out the essential feature of the experiment, the fact that agent's are walking under the influence of the 'elderly' stereotype.

So, in order to put to rest this kind of objection, I must show not only that 'walking at 4.2kph' is an automatic action, but also that 'walking under the influence of the stereotype' is an automatic action – and so that the latter is under the agent's guidance too.

In fact, there are some descriptions of the agent's behaviour, in leaving the room after the experiment, which are obviously under the agent's guidance: if one describes what the agent does as 'walking', that is under the agent's guidance: the agent can at any time stop walking, as she can start running (or crawling, for that matter).

Also, if one describes what the agent does under the description 'walking at 4.2 kph', that description too is under the agent's guidance: the agent can at any time increase or decrease her speed (suppose, for example, that the agent receives an emergency phone call, or that the fire alarm goes off, or that she starts to wonder whether she has left her bag in the experiment's room). Now the question is: does the agent also have guidance over her action under the description 'walking under the influence of the stereotype'?

That an agent who is under the influence of such stereotypes can change her behaviour has been shown in a experiment by Macrae (1998), modelled on Bargh's version:

subjects were primed with the stereotype of ‘helpfulness’, and then put in a situation in which they could have picked up a pen that the experimenter pretended to accidentally drop. Macrae’s results matched Bargh’s: subjects who had been primed with the ‘helpfulness’ stereotype tended to pick up the pen more often then subjects in the control group. But Macrae added an element: sometimes the pen was working fine, and sometimes it was leaking. And he found that when the pen was leaking, there was no registered effect of the ‘helpfulness’ stereotype: primed subjects no longer tended to help more often than control subjects.

Macrae’s experiment appears to prove an obvious point: that the fact that agents are subject to the stereotype, and behave accordingly to it, does not mean that agents cannot change their behaviour: so, again, in Bargh’s scenario, agents would have obviously picked up their speed if the fire alarm had gone off, for example. The interesting question, there, is whether primed agents who rushed to the exit would have been still slower than control agents who rushed to the exit: Macrae’s findings, which registered no effect of the ‘helpfulness’ stereotype in the leaking pen cases, appear to suggest that primed agents would not have in fact been slower than control agents had the fire alarm gone off.

So what subjects do in leaving the experiment’s room, under descriptions such as ‘walking at 4.2 kph’ and ‘walking under the influence of the stereotype’, is under their guidance. And since I have already shown that it also meets my conditions for automaticity, we can conclude that it is a case of automatic action.

Here it might be objected that I have only shown that, had the circumstances changed, or had the circumstances been different, the agent would have behaved differently. But that this does not show that, in the actual circumstances, namely when Bargh registered a speed of 4.2kph, agents had guidance. The possible fire alarm case (or the phone ringing), then, would not show that the agent had guidance over the actual case; because in the actual case, no fire alarm went off. What I need to show for guidance, then, would be that agents are able to change their speed, or to stop walking at will, independently of a change in circumstances such as the fire alarm going off.

I think I can do that: suppose that, while walking down the corridor, one of the subjects that Bargh had just primed with the 'elderly' stereotype feels a sudden rush of affection for her new born baby, which, for the first time, she has left at home with someone else. It would be very weird to think that, because she is under the influence of the 'elderly' stereotype, she could not run home to hug her baby (thereby increasing her speed). That shows, I think, that subjects can change their speed at will, independently of the circumstances.

Here, a defender of the previous objection might still want to reply that I have not shown that the intervention is independent of the circumstances because, indeed, having a sudden rush of affection for your new born baby is a change in circumstances. This did not happen in the actual case, and therefore it does not show guidance in the actual case.

First of all, we don't know that primed subjects did not have sudden rushes of affection in leaving the room. We only know that, if they had them, they did not result in increased speed (not even that, actually: 4.2kph might be the result of them walking faster than they would have had, had they not had the rush of affection). But, most importantly, if the defender of the objection is willing to reduce any act of will to a change of circumstances, then they would have set a target, a pure act of will, that they themselves consider unreachable. But then they would end up with the counterintuitive position of denying control not only in the 'elderly' stereotype cases, but in all cases; because any case could be potentially reduced to a change of circumstances.

But then their objection would no longer concern Bargh's cases; it would just be a general objection about the possibility of control. They would, indeed, be accepting my point that Bargh's cases, despite the stereotype, resemble normal cases; it's just that they would deny that normal cases are under the agent's guidance in the first place. But then this is not the place to take on their general scepticism about control (and maybe, eventually, free will).

Another objection might be that 'acting under the influence of the stereotype' might be helped, but that the agent cannot help *being* under the influence of the stereotype. This doesn't matter: it is with the agent's behaviour that we are concerned; and with whether the agent has guidance over her behaviour being affected by the stereotype. Once we have shown that she does, then it doesn't matter that she can't help *being* (whatever that means) under the influence of the stereotype.

So, we can conclude, the agent's behaviour, in leaving the experiment's room, is under the agent's guidance or intervention control under both 'walking at 4.2kph' and 'walking under the influence of the stereotype': so those are both cases of automatic action. In this section I have therefore shown that the cases discussed by Bargh are cases of 'automatic action' in my sense.

I can imagine the reader being rather disappointed with my discussion of Bargh. There was no mention, in Chapter 1, of automaticity being about psychologists influencing the behaviour of people. And, when faced with the frightening side of automaticity - some people's ability to make other people do as they wish - I just contented myself with demonstrating that those are cases of automatic action too. But there must be a salient difference, the reader will object, between someone being made to walk slower (or being made to buy one particular product rather than another, or being made to help someone rather than not, and, more worryingly, *vice versa*) through influences of which she isn't even conscious, and 'turning a door handle'. The former is scary; the latter is just 'turning a door handle'.

It was not my intention to disregard the distinctiveness of Bargh's cases (and, possibly, their social relevance); but *it was* my intention to normalize them. My concern, in arguing that Bargh's cases are actions, is to show that they do not imply any diminished responsibility or diminished control. I don't think, in short, that it makes sense to say that those agents that have been primed with the elderly stereotype have less control over the speed at which they walk down the corridor than those agents that have not been primed with the elderly stereotype. Yes, primed subjects walk slower than control subjects. And yes, primed subjects walk slower than control subjects because of the elderly stereotype. But I don't think that this amounts to any determination or diminished control, and therefore I don't think that their walking slower implies any diminished responsibility.

The point is that primed agents are as responsible for walking at 4.2 kph as control agents are responsible for walking at 4.8 kph. It's as if the first group of subjects, rather than being primed, would have to walk down a corridor with an acclivity so slight that they would not realize there was one; while the second group would walk down a perfectly flat corridor. Predictably, the first group would be on average slower than the second group; suppose, again, 4.2 kph against 4.8 kph. Would it make any sense to say that the first group is less responsible for doing 4.2 kph than the second group is for doing 4.8 kph? I don't think it would.[5]

But one might want to argue, rather, that the first group is not responsible for walking slower than the second group. And that, similarly, the primed group is not responsible for walking slower than the control group. But what does that mean? We know from

[5] It might be argued that the fact that, in the first case, part of the responsibility lies with someone – the experimenter – who is absent from the second case, must mean that subjects in the first case are less responsible than in the second because they share their responsibility with the experimenter. But responsibility does not work like that: it is not a cake. One's increased responsibility does not necessarily imply that someone else's responsibility must decrease. Suppose I was given life for having planted a bomb on a school bus. If it later came out that I had an accomplice, that would not decrease my responsibility – even though I would obviously share my responsibility with them.

both commonsense and Macrae's (1998) findings that subjects in the first group are, in both cases, perfectly able to increase their speed. And that suggests that neither the slight acclivity of the corridor nor the elderly stereotype could excuse the agent if she was, say, late in picking up her kids from school.

Contrast this case with the case in which the door at the end of the corridor is locked. The agent cannot leave, she is stuck. Now this is a case in which the agent might be excused from being late in picking up her kids. Crucially, it emerges that the agent's awareness of the bias doesn't matter: in the case of the locked door the agent is aware of the bias and still she is excused. In the case of the elderly stereotype the agent is not aware of the bias but nevertheless she is not excused.

2. Dual control

In this section I present the well-established empirical hypothesis of dual control, in the version by Norman and Shallice (1986[6]), showing that the behaviours for which it gives a causal explanation (explanation with which this thesis is not concerned) meet my necessary and sufficient conditions for automatic action.[7]

[6] A very similar version of dual control is presented by Perner (2003). Even though most of what I shall say about Norman and Shallice would apply to Perner's model, here I shall not concern myself with it.

[7] I don't think it can be the job of a philosopher to provide empirical evidence for an empirical thesis. Therefore, anyone who's interested in the empirical evidence for dual control, should look at the following: Norman and Shallice (1980), Shallice (1982), Norman and Shallice (1986), Shallice (1988), Shallice and Burgess (1996), Cooper and Shallice (2000).

The basic idea of dual-control is that there is a conscious level of control and a non-conscious automatic level of control.[8] And that some performances can be carried out by the lower level of control without the involvement of the conscious level, while other performances require the supervision of consciousness. In this section I will be showing that those performances that can run without the involvement of the conscious level of control meet my conditions for automatic action.

After briefly sketching Norman and Shallice's model, I shall argue that the behaviours the model proposes to explain are not only *similar* to automatic actions, they actually *are* automatic actions – just because these behaviours meet the necessary and sufficient conditions for automatic action from Chapter 1.

This is, firstly, how Norman and Shallice describe their model's goals:

> Our goal is to account for several phenomena in the control of action, including the several varieties of action performance that can be classified as automatic, the fact that action sequences that normally are performed automatically can be carried out under deliberate conscious control when desired, and the way that such deliberate control can be used both to suppress unwanted actions and to enhance wanted ones. In addition, we take note both of the fact that accurate, precise timing is often required for skilled performance and the fact that it is commonly believed that conscious attention to this aspect of performance can disrupt the action (Norman and Shallice 1986, p. 3).

[8] I will here and throughout refrain from speaking of "automatic control" because that name also refers to the engineering discipline which studies systems such as thermostats (I know, philosophy studies thermostats too!). Perner (2003) talks of *vehicle control* for the lower level and *content control* for the higher level.

This shows that Norman and Shallice share much of my motivation for discussing automatic actions. First, they emphasize that “several” different kinds of actions can be classified as automatic. Indeed, in Chapter 1 I have shown how performances as different as skilled actions and habitual actions can both be said to be automatic; and that there is a difference between spontaneous and subsidiary automatic actions. Furthermore, Norman and Shallice emphasize, as I do, that an action *becomes* automatic through practice; and that (possibly because of that) the fact that a performance has become automatic does not mean that it is now beyond the agent’s attention and conscious control.

Also, the agent can draw her attention to an automatic performance and consciously control such a performance “when desired”. Indeed, an agent can intervene both to “suppress” unwanted aspects of a performance and to “enhance” more appropriate ones. Finally, Norman and Shallice, like me, want to account for the “commonly believed” intuition that attention, in the case of automatic action, can “disrupt” the performance.

Here are some further similarities between what Norman and Shallice want to account for and what I say about automatic actions in Chapter 1:

> The theory must account for the ability of some action sequences to run themselves off automatically, without conscious control or attentional resources, yet to be modulated by deliberate conscious control when necessary. Accordingly, we suggest that two complementary processes operate in the selection and control of action. One is sufficient for relatively simple or well learned acts. The other allows for conscious,

> attentional control to modulate the performance (Norman and Shallice 1986, p. 1).

Norman and Shallice's references to both "well learned acts" and to the intervention of "conscious control when necessary" already point to my concept of automatic action.

On Norman and Shallice's model, the lower level of control is regulated by what they call *contention scheduling*: action schemas - potential reactions to some perceptual input - have different activation values in relation to some perceptual input; and a perceptual input will activate an action schema if that action schema has a low enough activation value in relation to that perceptual input, such that contention scheduling can select that action schema without the intervention of consciousness.

Here's an example: my office door doesn't have a handle; I just push it to come in. So, supposedly, when I'm accessing my office, the activation value of the action schema 'push the door' will be lower than the activation value of the action schema 'turn the door handle'. Also, because I go in and out of my office dozens of times a day, the action schema 'push the door' when I receive the perceptual input of seeing my office door from the outside will supposedly be low enough to be selected by contention scheduling, so that I will often push my door open automatically.

> There are two basic principles of the contention scheduling mechanism: first, the sets of potential source schemas compete with one another in the determination of their activation value; second, the selection takes place on the basis of activation value alone – a schema is selected whenever its activation exceeds the threshold that can be specific to the

> schema and could become lower with use of the schema (Norman and Shallice 1986, p. 5).

Sometimes, according to Norman and Shallice, some perceptual input is such that no action schema has, in relation to it, an activation value low enough for the schema to be selected. Think, for example, of novel experiences; or things one is not very good at. Those reactions to the environment cannot be dealt with by contention scheduling alone: the intervention of consciousness is required.

> We propose that an additional system, the Supervisory Attentional System, provides one source of control upon the selection of schemas, but it operates entirely through the application of extra activation and inhibition to schemas in order to bias their selection by the contention-scheduling mechanisms (Norman and Shallice 1986, p. 6).

There are, then, two functions for the Supervisory Attentional System (SAS): it lowers down activation values when no action schema has a low enough value to be selected by contention scheduling; and it inhibits action schemas that, despite their activation value being low enough to be selected by contention scheduling, are inappropriate to the circumstances.

An example of the latter is when one is involved in a familiar activity, but this time the agent has to do something slightly different. Think of driving on familiar roads towards unfamiliar destinations: I usually go this way on my way home, but today I am headed to a friend's house. Being on a familiar road, the action schemas relevant to driving home have very low activation values, sufficient to be selected by contention scheduling – just because I have gone this way endless times. But this time I am going somewhere else: therefore the SAS must inhibit selection of the inappropriate action schemas.

3. Dual control and automatic actions

Having described how Norman and Shallice's dual control model is supposed to work, I will now show, by looking at their writings, that the behaviours the model is supposed to explain meet my necessary and sufficient conditions for automatic actions.

The first thing to look at is what Norman and Shallice mean by the term 'automatic':

> Examine the term automatic… First, it refers to the way that certain tasks can be executed without awareness of their performance (as in walking along a short stretch of flat, safe ground). Second, it refers to the way an action may be initiated without deliberate attention and awareness (as in beginning to drink from a glass when in conversation) (Norman and Shallice 1986, pp. 1-2).

The first point is one that I have also made in Chapter 1: there is no need for awareness in order to execute automatic actions. This is, given that on my account awareness depends on attention, my second necessary condition on automaticity: no need for attention. The example is particularly illustrative: when walking, an agent need not be aware of her legs' movements in order for her legs to perform.

The second point makes, again, specific reference to the fact that attention and awareness are not necessary: but in this case Norman and Shallice talk of the initiation of the performance rather than its execution. I have not drawn such a distinction in

Chapter 1, but it appears clear that, on my account, attention and awareness are not necessary, and are lacking, both in the initiation and in the execution of the performance. Because if one's attention was caught or employed for either initiation or execution, then an agent would be attending to the performance, and then the performance would no longer be automatic.

Here the reader might accept that the relationship between attention and awareness on the one side and automaticity on the other envisaged by Norman and Shallice and the one I establish in Chapter 1 is the same; but they might question whether Norman and Shallice might mean something else by 'attention' and 'awareness'. After all, they aren't philosophers. And so the fact that they use the same words is no guarantee for the fact that they are talking about the same phenomena.

In answering this point, I should first clarify something: in order to show what I want to show in this section, namely that dual control models are an example of the relevance of automatic actions in empirical literature, I don't actually need to show that Norman and Shallice mean exactly what I mean by attention, awareness, or automaticity. Indeed, their being scientists rather than philosophers, it is hard to imagine that they could ever mean the very same things by those terms. It is sufficient, for this section, that I show that those behaviours they are out to explain meet my conditions for automatic action.

Having said that, could their conception of attention and awareness be incompatible with mine? A suggestion that it might be comes from another passage:

> It is possible to be aware of performing an action without paying active, directed attention to it. The most general situation of this type is in the initiation of routine actions (Norman and Shallice 1986, p. 2).

Here Norman and Shallice deny that lack of attention implies lack of awareness: they propose, indeed, that one can be aware of some performance without paying "active, directed" attention to it. I have not distinguished between active and passive attention, or between directed and non-directed attention. I have said, though, that what I mean by an agent attending to some action includes the case in which the agent's attention is caught by the action rather than being directed to the action. Furthermore, I have said that both ways of attending are ways of becoming aware.

Therefore, if Norman and Shallice's position, as the text suggests, is that one need not direct one's attention to some performance in order to become aware of it, then my account from Chapter 1 makes that point too: one can also become aware by one's attention being caught by the performance.

But there is another possible inconsistency that the above passage from Norman and Shallice might suggest: namely, that routine actions do not imply lack of awareness; and eventually, even though Norman and Shallice don't specifically say that, that automatic performances might not require lack of awareness. I have denied that point in Chapter 1: the first necessary condition on automaticity is lack of attention and awareness.

But I think that this suggestion is not grounded in Norman and Shallice's text, as can be seen from the passage below (which directly follows the last one I quoted):

> Phenomenally, this corresponds to the state that Ach (1905) describes as occurring after practice in reaction time tasks. Over the first few trials, he said, the response is preceded by awareness that the action should be made, but later there is no such awareness unless preparation has been inadequate (Norman and Shallice 1986, p. 2).

With practice, then, awareness disappears; and we find awareness only if we have not practiced enough. On my account, it is only after practice that performances become automatic. So the idea that, to begin with, the agent is aware of her performances is not incompatible with my account; because, to begin with, the agent's performances are not automatic.

So far, then, the behaviours that the empirical hypothesis of dual control is concerned with meet both my individually necessary and jointly sufficient conditions for automaticity: lack of attention and awareness; and no need for attention and awareness.

Another suggestion of the fact that the behaviours discussed by Norman and Shallice are automatic in just the sense I have individuated in Chapter 1 comes from their list of activities that are not automatic:

- They involve planning or decision making
- They involve components of troubleshooting
- They are ill-learned or contain novel sequences of actions
- They are judged to be dangerous or technically difficult

- They require the overcoming of a strong habitual response or resisting temptation (Norman and Shallice 1986, pp. 2 and 3).

Now look at my examples of non-automatic actions from Chapter 1:

> ...finally confessing to a wrong-doing; holding on to the rope from which your best friend is hanging; driving through a snow-storm on a mountain road, at night (Chapter 1, p. 17).

The similarity is striking. My first example (confessing) clearly belongs to category 1. My second example belongs to both 3 and 4. My third example belongs to both 2 and 4. Finally, recall that in Chapter 1 I have said that many automatic actions can be found in habits and skills: indeed, category 5 explicitly refers to habits ("habitual response"); and category 3 makes clear reference to skills ("ill-learned").

Clearly, what I have established so far does not yet show that the behaviours that Norman and Shallice discuss are automatic actions; I still have to show that they meet my third necessary condition on automatic action, guidance. It does nevertheless show that we share a conception of automaticity.

Even this point, though, must be clarified: as I said before, it would be surprising if our concept of automaticity was exactly the same, given that theirs is based on empirical work and considerations, while mine is the result of conceptual analysis. So, in that sense, it would be pointless to try and argue that our concepts of automaticity are identical. Nevertheless, there is a sense in which they are: they refer to the same set of behavioural performances. And this sense is all I need in order to make the point that those behaviours are automatic in my sense.

4. Dual control and Guidance

Having established that Norman and Shallice's behavioural performances are automatic in the sense I individuated in Chapter 1, in this section I will argue that, meeting my guidance condition, they are also actions; and that therefore they are automatic actions.

Let us remind ourselves of what guidance is: it is the agent's ability to intervene to stop herself from doing something. And it is only those performances over which the agent has guidance that, I argue in Chapter 1, can be said to be actions. So are the performances that Norman and Shallice want to explain through contention scheduling under the agent's guidance?

Norman and Shallice say early on in their article what they think the relationship between automatic performances and control is:

> Our goal in this chapter is to account for the role of attention in action, both when performance is automatic and when it is under deliberate conscious control (Norman and Shallice 1986, p. 1).

According to Norman and Shallice, then, automatic performances are not under deliberate conscious control. And this is a point I have also made in Chapter 1, when I said that lack of attention and awareness implies lack of conscious control. But this is not incompatible with guidance: indeed, in Chapter 1 I show that an agent can have guidance over some performance that she is not consciously controlling.

It might be thought that the incompatibility arises because of contention scheduling. Recall that if some action schema has low enough activation value relative to a perceptual input, it will be selected. It might be thought that, because this process happens without the agent's awareness, the agent cannot stop a low enough action schema from being selected. And that this means that the agent has no guidance over the selection of that action schema. If that was the case, then the performance resulting from the selection of that action schema could not be a case of automatic action, because it would fail to meet one of my necessary conditions.

Suppose that I drive from a country where green means 'go', to a country where green means 'stop'; and suppose that I have never before been in a country where green means 'stop'. Now, we can suppose that I am so used, from decades of driving, to press on the accelerator when I see green that, according to Norman and Shallice, 'pressing on the accelerator' has very low activation value in relation to perceptual input 'green'; low enough, indeed, that it will be selected by contention scheduling without the involvement of the SAS.

Having shown that performances that are selected by contention scheduling are automatic in my sense, we can say that I often automatically press on the accelerator when the lights turn green. Does this mean that I have no guidance over 'pressing on the accelerator'? I don't think so: when I drive across the border, I will have to remember that 'green' no longer means 'go'. But it won't be a case of being unable to stop myself from pressing on the accelerator when I see green. It will only be a case of paying more attention than usual, so that my reaction to green won't be, this time, automatic.

And I think that Norman and Shallice acknowledge this point too:

> …a schema might not be available that can achieve control of the desired behaviour, especially when the task is novel or complex. We propose that an additional system, the Supervisory Attentional System (SAS), provides one source of control upon the selection of schemas, but it operates entirely through the application of extra activation and inhibition to schemas in order to bias their selection by the contention scheduling mechanisms (Norman and Shallice 1986, p. 6).

Once in the foreign country, then, the task of reacting to 'green' will be novel; and complex, indeed, exactly because I must make sure not to slip back into the old habit. Because of this, according to Norman and Shallice, the task cannot be dealt with by contention scheduling alone. There are, indeed, two issues, and Norman and Shallice cover them both: activation of the novel reaction, 'stopping when the light is green'; and inhibition of the habitual reaction, 'pressing on the accelerator when the light is green'. Both activating and inhibiting are, here, too much for contention scheduling alone.

But this, Norman and Shallice acknowledge, does not mean that the agent has no control over her reactions to 'seeing the green light'; but only that the agent must deal with 'seeing the green light' differently. In this case the agent will have to attend to what she does, as Norman and Shallice acknowledge: “Attention, which we will associate with outputs from SAS…” (Norman and Shallice 1986, p. 7).

So not only are behaviours selected by contention scheduling not incompatible with guidance, but actually Norman and Shallice too defend the idea that agents can intervene to correct, guide, or inhibit their automatisms.

In conclusion, then, the behaviours that the hypothesis of dual control is supposed to explain not only meet my conditions for automaticity, but they are also under the agent’s guidance; they are, therefore, automatic actions.

Conclusion

In this chapter I have looked at two very influential examples of the importance of automatic actions within empirical psychology: Bargh’s automatic influences and Norman and Shallice’s dual control model. I have shown that Bargh’s cases, such as subjects walking slower out of a room where they have just been primed with the ‘elderly’ stereotype, meet my necessary and sufficient conditions for automatic action. Then I have presented the hypothesis of dual control, according to which there are an automatic unconscious level of control, and a conscious level of control. I have shown that those behaviours that, according to Norman and Shallice, are controlled by the lower level of control, are automatic. And, since they are also under the agent’s guidance, they are in fact automatic actions. So, not only has this chapter shown how psychology is interested in automatic actions; it has also shown that psychologists accept the definition of automatic action that I have given in Chapter 1.

I now consider the task of presenting automatic actions concluded. I will therefore move on to establish, in Chapters 3 and 4, whether causal accounts of intentional action, such as Davidson's reductive one (Chapter 3) and Bratman's non-reductive one (Chapter 4) can account for the intuition that automatic actions are intentional.

<u>Chapter 3</u>: Davidson, Unconscious Beliefs, and Causes

Introduction

In this chapter I discuss Davidson's account of intentional action, as presented in *Actions, Reasons, and Causes* (1963). I show that Davidson's account does not work for automatic actions. This is because Davidson's account relies on the attribution of particular mental states as the causes of action in every case. But, I argue, there is no evidence for thinking that those mental states are *always* present in *all* automatic cases; nor is there evidence for thinking that they are always the causes of automatic actions. Furthermore, since those mental states, in automatic cases, must always be unconscious, they can always be attributed consequentially: but then, I argue, those mental states lose explanatory power.

The chapter comprises five parts: in the first, I present Davidson's view. In the second, I show that, for automatic cases, Davidson's view necessarily needs to appeal to unconscious mental states. In the third I show that, in automatic cases, there is no evidence for attributing in every case the unconscious mental states required by Davidson's view. In the fourth section I present an argument against the attribution of those unconscious mental states: the argument from consequential attribution. Finally, in the fifth section I deal with a possible reply from the Davidsonian camp.

1. Davidson's view

Davidson’s account of intentional action is put forward in his famous article *Actions,* *Reasons, and Causes* (1963), where Davidson defends the thesis that reasons explanation (rationalization) is "a species of causal explanation" (ibid, p. 3).

Let me first say which part of Davidson's argument I am interested in here: I am not concerned with Davidson's main contention that rational explanation (rationalization) is a form of causal explanation. I will rather focus my attention only on what emerges from *Actions, Reasons, and Causes* as Davidson's account of intentional action – see below. It is fair to say that an account of intentional action was probably not Davidson's main concern in writing *Actions, Reasons, and Causes*. But in this thesis *I* am only after one such account.

Furthermore, I should make the rather obvious point – at least for those who have read *Actions, Reasons, and Causes* – that Davidson does not speak of automatic actions. So the aim of this chapter is not to analyse Davidson's application of his account of intentional action to automatic actions. It is rather to establish whether Davidson's account of intentional action, which was not developed for nor applied to automatic actions, can be applied to them.

On Davidson's account, then, some action A is intentional under description φ only if that action was caused by a primary reason of the agent comprising of a pro attitude towards actions with a certain property, and a belief that action A, under description φ, has that property[1]:

[1] Davidson only offers necessary conditions. Any attempt at giving sufficient conditions would, by Davidson's own admission (Davidson 1973), run against the problem of deviant causal chains (more on this in Chapter 5).

> R is a primary reason why an agent performed the action A, under description d, only if R consists of a pro-attitude of the agent towards actions with a certain property, and a belief of the agent that A, under the description d, has that property (ibid, p.5).

Pro attitudes, says Davidson, can be “desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values” (Davidson 1963, p. 3).

So, on Davidson's account, my flipping the switch is intentional under the description ‘flipping the switch’ only if it was caused by a primary reason composed of a pro attitude of mine towards actions with a certain property, say the property of ‘illuminating the room’; and a belief that my action, under the description ‘flipping the switch’, has the relevant property of ‘illuminating the room’.

The crucial element of Davidson’s view is that the primary reason, composed of a pro attitude plus a belief, is the action’s cause. As Davidson himself points out (ibid, p. 12), causes must be events, but pro attitudes and beliefs are states, and so they cannot be causes. Davidson therefore proposes the “onslaught” (or *onset*, see Lowe 1999, p. 1) of the relevant mental state as the cause of action.

The difference between a mental state and its onset, which is a mental event, is the same as the difference between believing that there is a bottle on my desk (mental state), and forming the belief (noticing, realizing) that there is a bottle on my desk (mental event). Clearly, while both mental states, pro attitude and belief, are always needed to rationalize an action under some description, only one mental event is necessary to cause the action.

As Stoutland (1985) emphasizes, the mental states required by Davidson's view must have a very specific content:

> The thesis is a very strong one: it is not saying merely that reasons are causes of behaviour but that an item of behaviour performed for a reason is not intentional under a description unless it is caused by just those reasons whose descriptions yield the description under which the behaviour is intentional. This requires that every item of intentional behaviour have just the right cause (Stoutland 1985, p. 46).

So there must be a content relation between the primary reason and the action description in question. Recall Davidson's definition of "primary reason" (Davidson 1963, p. 5): the belief must make explicit reference to the action description which it rationalizes.

The following primary reason, for example, would not do: a pro attitude towards 'illuminating the room', and a belief that my action, under description 'turning on the light', has the property of 'illuminating the room'. This primary reason makes no mention of the description 'flipping the switch', and therefore it cannot rationalize my action under the description 'flipping the switch'. Even though it will rationalize my action under the description 'turning on the light'.

One note of clarification: the content constraint emphasized by Stoutland is on the belief rather than on the pro attitude. That is to say that, as long as the belief has the 'right' content, the pro attitude can have any content. For example, my action of flipping the switch can be rationalized under the description ‘flipping the switch’ by a very wide selection of pro attitudes - ‘turning on the light’, ‘illuminating the room’, ‘wasting energy’, ‘finding some comfort’, ‘stretching my arm’, etc. – as long as the agent believes that her action, under the description in question – ‘flipping the switch’ – has the relevant property towards which the agent has a pro attitude: ‘turning on the light’, say.

A peculiar case will be represented by the case in which I flip the switch with a pro attitude towards ‘flipping the switch’. In this case, the content of the belief will be tautological: that my action, under description 'flipping the switch’, has the property of 'flipping the switch' (I return to these sorts of cases in Section 3.1.1).

1.1 Inference to the best explanation

Before arguing against Davidson's view, I must clarify what Davidson takes to be the nature of his argument. This is particularly important since in this chapter I will be arguing, primarily, that there are no arguments in favour of the application of Davidson’s view to automatic actions, rather than arguing for the impossibility or incoherence of such application.

Davidson admits that he has no positive argument in favour of his causal view:

> ...failing a satisfactory alternative, the best argument for a scheme like Aristotle's [a causal account] is that it alone promises to give an account for the 'mysterious connection' between reasons and actions (Davidson 1963, p. 11).

If this were to apply also to Davidson's account of intentional action, then the reason for thinking that an action is intentional only if it is rationalized by a primary reason *which is its cause* would simply be that there is no "satisfactory alternative": therefore Davidson's argument wholly relies on this assumption about the absence of a "satisfactory alternative".

Indeed, Davidson's argument could be taken to be an inference to the best explanation.[2] We have a case of 'inference to the best explanation', in the words of Harman (1965, the first to use this expression), when "one infers, from the premise that a given hypothesis would provide a 'better' explanation for the evidence than would any other hypothesis, to the conclusion that the given hypothesis is true" (Harman 1965, p. 89).

Under this comparative understanding of the value of an hypothesis, then, to argue against some theory one must show that there is a better view. In the absence of such a view, the hypothesis under scrutiny must be taken to be true (as long as it is, it should be added, consistent). And what Davidson says is indeed that there is no satisfactory alternative to his view. So, on this understanding, rather than arguing against Davidson's view, in this chapter I should look for alternative views of intentional action.

I am willing to accept this point, because in Chapter 5 I will develop an alternative account of the intentional character of automatic actions. So if one wants to

[2] I owe this point to Tony Booth.

understand Davidson's argument as an inference to the best explanation, then one should take what I say in this chapter against Davidson's view, and what I say in Chapter 5 in favour of my own account, as reasons for thinking that my own account is a better view of automatic actions than Davidson's.

Obviously, since here I am only interested in the application of Davidson’s view of intentional action to automatic actions, it is only on those grounds that Davidson’s view must be compared to the one that I present in Chapter 5. Claims such as Davidson’s contention that rationalization is a form of causal explanation, for example, should play no part in the comparative assessment of Davidson’s view and mine.

1.2 Shortcut

There is an obvious shortcut that one could take in arguing against Davidson's view. As the reader will recall from Chapter 1, a necessary condition on automatic action is lack of attention and awareness. Therefore, if Davidson's view required that the agent be aware of her actions, then Davidson's view couldn't work in the case of automatic actions. In this section I will show that this shortcut isn't available, because it is not possible to establish, from Davidson's writings, whether awareness of action is indeed a requirement of his account.

Firstly, though, I must clarify the relationship between the agent's awareness of her actions and the agent's awareness of her reasons, since the two passages from Davidson that I will be looking at concern the agent’s awareness of actions and awareness of reasons respectively. Only the former is a necessary condition on automatic action, so that an action can be automatic, under some description, only if the agent isn't aware of it under that description (see Chapter 1). But if an agent is aware of a primary reason of hers, then she will be aware of her action under the description which is rationalized by the primary reason of which the agent is aware. This is simply because, as we have seen in Section 1, it is a requirement of Davidson's thesis that the action description be part of the content of the primary reason: specifically, of its belief component.

Therefore an agent can be unaware of some action description only if she is not aware of the primary reason (or just its belief component) which rationalizes that action description. The agent's awareness of the primary reason, then, given the content constraint (see Section 1), is sufficient for the agent's awareness of her action under the description being rationalized. Specifically, since it is the content of the belief that must refer to the action description, the agent mustn't be aware of the belief, if her action is to be automatic under the description which that belief rationalizes. The agent needn't be unaware of the pro attitude, since, as we have already seen, the pro attitude doesn't need to make reference to the action description being rationalized. But if the agent was aware of the whole primary reason, then the agent would be aware of the belief component. So if the primary reason is understood as a whole, then the agent mustn't be aware of it, because the agent mustn't be aware of its belief component.

Suppose, for example, that I flip the switch because I had a desire to turn off the light and a belief that my action, under description 'flipping the switch', had the property of turning off the light. Since 'flipping the switch' is part of the content of the agent's belief, if the agent is aware of her belief, then the agent is aware of 'flipping the switch'.

Here I am therefore ruling out the possibility of an agent who is aware of her pro attitude towards actions with the property P, and is aware of her belief that her action, under description φ, has property P, but is unaware of φ-ing. But one could devise odd cases in which an agent who is aware of the relevant primary reason falls unconscious on the point of acting, but still has the luck of completing her action, of which she would therefore be unaware, in a way in which the action satisfies the primary reason.

One could suppose, again, that I had a pro attitude towards turning off the light, and a belief that my action, under description 'flipping the switch', had the property of 'turning off the light'; and that I was aware of my belief. But that, on the point of flipping the switch, I fell unconscious. Nevertheless, by a stroke of luck, my hand falls upon the switch and flips it anyway, but I can't be aware of it because I have fallen unconscious. Now this might be a case in which I am aware of my reasons without being aware of my actions; but, intuitively, given the crucial role of luck and the obvious absence of control, this doesn't even look like an action of mine, never mind an intentional action. So I don't need to worry about these sorts of odd cases. I am happy to restrict what I said above about the relationship between awareness of reasons and awareness of actions to actions that are intuitively intentional.

Having clarified the relationship between awareness of actions and awareness of reasons, the question is whether the former is a requirement on Davidson's view. But also, as we have just seen, were the latter to be a requirement on Davidson's view, the view would violate the lack of awareness condition on automatic action. There are two places in Davidson's writings to which we might refer in order to answer this question. Below is the first:

> To dignify a driver's awareness that his turn has come by calling it an experience, or even a feeling, is no doubt exaggerated, but whether it deserves a name or not, it had better be the reason why he raises his arm (Davidson 1963, pp. 12-13).

Here, despite saying that it would be "exaggerated" to describe the driver as having an 'experience' or 'feeling', Davidson not only speaks of the "driver's awareness", but he says that the driver's awareness of the turn had "better be the reason why he raises his arm"; implying that, had the driver not been aware of the turn, his behaviour couldn't have been rationalized. If that were Davidson's meaning, then automatic actions, the agent having to be unaware of them, could not, on Davidson's view, be rationalized.

But in a later article, when discussing awareness of reasons, Davidson appears to take a position that is, for our purposes, importantly different:

> We cannot suppose that whenever an agent acts intentionally he goes through a process of deliberation or reasoning, marshals evidence and principles, and draws conclusions. Nevertheless, if someone acts with an intention, he must have attitudes and beliefs from which, had he

> been aware of them and had the time, he could have reasoned that his action was desirable (or had some other positive attribute) (Davidson 1978, p. 85).

Here it appears clear that the attitudes and beliefs which compose primary reasons need not be mental states of which the agent is aware at the time of action. Indeed, Davidson says "had he been aware of them", which must imply that the agent wasn't aware of them; and, therefore, that agents need not be aware of the mental states which rationalize their actions; so the action descriptions rationalized by those mental states can be automatic (if they fulfil my other criteria).

Even though there is some discrepancy between Davidson's views in the different articles, I shall conclude – being as charitable to Davidson as possible - that there isn't enough evidence from Davidson's writings to conclude that awareness of action is indeed a requirement on his view. So, at least on that ground, there is no incompatibility in principle between Davidson's account of intentional actions and automatic actions as I have defined them in Chapter 1.

In the next section I show that, in automatic cases, the beliefs needed by Davidson's view have to be unconscious, given the lack of awareness condition on automatic actions.

2. Unconscious mental states

As I have just shown, if Davidson's theory is to have any chance of being applied to automatic actions, then the agent mustn't be aware of her primary reasons. Again, it must be specified that it isn't the whole primary reason that has to be unaware or unconscious in the automatic case, but only its belief component; the one that, on Davidson's account, must make reference to the action description.

So, if Davidson's theory is to have any chance of working for automatic actions, it must appeal to unconscious mental states: as in, mental states of which the agent is not aware. In the literature (see Searle 1992) there is a distinction between two kinds of unconscious states: unconscious states such as my belief that 'the Eiffel Tower is in France', and unconscious states such as 'the myelination of the axons in the nervous system'. The former, says Searle, is unconscious most of the time (not now) because we almost never entertain the belief that 'the Eiffel Tower is in France' – obviously, when we do entertain such belief, then it isn't unconscious. Nevertheless, such belief, when it is unconscious, is still accessible by consciousness.

The latter kind of state, on the other hand, is unconscious just because it is not the right kind of state for consciousness to access or entertain – it is, in Dennett's (1969) terminology, a sub-personal state, while the belief that 'the Eiffel Tower is in France' is a personal state, even when it is unconscious, because it is accessible. Searle proposes to call the 'Eiffel Tower is in France' kind of state *unconscious state*, and the 'myelination of the axons in the nervous system' kind of state *non-conscious state*: I will stick to Searle's terminology.

It is important as much as obvious to point out that it is only the unconscious kind of state that can be deployed by Davidson's theory in accounting for automatic actions. The latter kind, the non-conscious state, being a subpersonal state of the brain, does not have content because it is not *about* anything; and it therefore cannot rationalize behaviour, given that, as we have seen, according to Davidson there must be a content-relation between the belief component and the action description being rationalized.[3]

So the kind of unaware belief that Davidson’s theory needs in the automatic case is what Searle calls an unconscious mental state: a mental state which is unconscious because the agent is not aware of it, but that can be accessed by the agent: it can be 'directly' brought to consciousness (as opposed to, say, discovering one's own states of mind by looking at a brain scan (which works for the non-conscious kind too), or by going to the psychoanalyst[4]).

So, if Davidson’s theory is to rationalize E’s automatically flipping the switch, it has to attribute to the agent E some pro attitude towards, say, ‘saving energy’, plus an unconscious belief that E’s action, under the description ‘flipping the switch’, has the property of ‘saving energy’. What it means for such belief to be unconscious is, according to Searle (1992), that the agent isn’t aware of her belief at the time, but that the belief is, nevertheless, accessible.

[3] Here one could appeal to the distinction between propositional and non-propositional content (see Crane 2003, p. 1), and argue that subpersonal states can at least have content in the latter sense. It is not at all clear that subpersonal states can have non-propositional content either (which is usually rather applied to mental states other than propositional attitudes, such as emotions); but even people, such as Bermudez (1995), who claim that subpersonal states can have content accept that subpersonal explanation isn’t rational explanation. Therefore even from their point of view a subpersonal state cannot rationalize behaviour.

[4] The kinds of states of mind that can be discovered through psychoanalysis must be therefore distinguished from Searle's unconscious states, because they are accessible in a different, external, way. Moran (2001) and Romdenh-Romluc (forthcoming) divide what Searle calls unconscious states in *subconscious states* (the ‘Eiffel Tower’ kind) and *unconscious states* (the Freudian kind).

This picture does, indeed, suit intuitions about automatic actions: when we do something automatically, such as flipping a switch, we do not think about it, we do not deliberate, we do not pay attention to what we do. But that doesn't necessarily mean, supporters of Davidson's picture will want to say, that we don't have the relevant beliefs. It is just that those beliefs are unconscious. And as long as those mental states are unconscious, Davidson's picture does not clash with our intuitions; nor, as I have shown, does it clash with my definition of automatic actions from Chapter 1.

But that it is possible for those unconscious beliefs to be attributed, so that Davidson's view can work for automatic actions, is still no argument in favour of their attribution. In the next section I will argue that, in automatic cases, there is not always evidence for the attribution of the required unconscious belief. Therefore Davidson's account does not work for all cases; so that his necessary condition – that actions are intentional *only if* they are rationalized by a primary reason – does not stand; because amongst intuitively intentional actions there are some automatic ones that Davidson's account fails to rationalize: so that being rationalizable isn't necessary for being intentional.

Before proceeding with my argument, though, I must deal with another potential objection: that I am setting my target too low. Rather than arguing against the attribution of unconscious mental states in all automatic cases, I should argue that it is never possible to attribute the required mental states in automatic cases.

Given that I am conceding to Davidson the possibility of the attribution, then, in the absence of a 'satisfactory alternative', Davidson's view should be chosen, because it can deliver, at least in principle, on the intentionality of automatic actions. But if Davidson's view is not the only one which can, in principle, deliver on the intentionality of automatic actions then, given the alternative (that I present in Chapter 5), it is not enough that attributing unconscious mental states is *possible*: we actually need an argument for doing so.

3. Arguments for the attribution

In this section I argue that there is not always evidence for thinking that in all automatic cases the agent has the relevant unconscious beliefs; and that those unconscious beliefs cause action. Then I look at two more possible arguments in favour of the attribution of the required unconscious beliefs, and I find that neither is conclusive.

3.1 Attributing unconscious beliefs

We have so far established that for Davidson's account to work for automatic actions, it may be possible to appeal to unconscious mental states as their rationalizers and causes. Now we need to find out whether there is any evidence for thinking that, in every automatic case, there indeed is an unconscious belief (or, as a whole, an unconscious primary reason) that causes and rationalizes each automatic action. The point here is twofold: each time, the agent must have had the relevant unconscious belief, and that unconscious belief must have been the cause of the agent's automatic action.

First of all, since the beliefs required by Davidson's thesis are unconscious, clearly Davidson's thesis cannot help itself to the most obvious ground for the attribution of mental states: that the agent has entertained the mental state; or, as Nagel (1970) famously put it, that there is something that it is like for the agent to be in that mental state. Unconscious mental states are, in this sense, *phenomenologically silent*. And therefore phenomenology does not constitute a reason to attribute unconscious mental states.

This point must be distinguished from the claim that phenomenology constitutes a reason against the attribution of unconscious mental states. The latter claim would be unfair: if a mental state is unconscious, then one cannot expect phenomenology to warrant its attribution. Phenomenology is just the wrong sort of domain.

One might think that there is a further reason why phenomenology is the wrong sort of domain: namely, that introspection isn't the right kind of evidence (as Wittgenstein (1953: § 551, 587, 591) appeared to think). If that were true, then phenomenology wouldn't just be the wrong sort of ground for unconscious mental states. It would actually always be the wrong sort of ground for the attribution of a mental state, no matter if conscious or not.

But, as I said, it is not with phenomenology that I shall concern myself with here. Another obvious ground for the attribution of the required beliefs is wanting to make sense of the agent's behaviour. If one flips a switch with the desire to illuminate the room, her behaviour makes sense only if she has the relevant belief - that flipping the switch will illuminate the room. But clearly this is not a sufficient ground for attributing the belief, otherwise we would have to accept that agents always do sensible things. So if we want to allow for the fact that agents act sometimes irrationally, we can't accept the above as sufficient grounds for the attribution of the required beliefs.[5]

Rather, I shall look into what appears to be the strongest ground in favour of Davidson's picture: the idea that agents, if asked for explanation (if asked what Anscombe (1957) called the 'why? question'), would answer by self-attributing a primary reason, or, anyhow, by self-attributing something that would enable us to construct a primary reason. And, more importantly for our purposes, that agents would do that even in automatic cases. So, for example, if E automatically flips the switch, then E, at the time of action, was unaware of flipping the switch. But the idea is that, if you pointed out to E her switch-flipping, and asked her why she had flipped the switch, she would answer with something quite similar to Davidson's primary reason: “I wanted to save energy”, say – which supposedly implies the belief that her action, under description 'flipping the switch', had the property of 'saving energy'. Are those self-attributions, then, evidence for thinking that the unconscious beliefs required by Davidson's thesis always cause automatic actions?

[5] Indeed, this is a further problem for Davidson's picture. Intuitively, when we flip a switch we normally act rationally. But if Davidson constrains our acting rationally to the agent having the relevant primary reason, and if I will be successful in showing that there are cases in which we don't have arguments for the attribution of the required belief, then Davidson will end up with cases in which agents intuitively act rationally but still his theory doesn't have grounds for the attribution of the beliefs that it requires in order to claim that the agents' behaviour was indeed rational. Just as with intentionality, I am pointing to cases that are intuitively intentional and intuitively rational – such as a normal switch-flipping, and then asking whether Davidson's theory can account for their intentionality and rationality. But here I will leave the ‘rationality’ part of my argument aside: it is only with intentionality that this thesis is concerned.

Note that Davidson allows for incomplete statements of reasons:

> A primary reason consists of a belief and an attitude, but it is generally otiose to mention both. If you tell me you are easing the jib because you think that will stop the main from backing, I don't need to be told that you want to stop the main from backing... Similarly, many explanations of actions in terms of reasons that are not primary do not require mention of the primary reason to complete the story... Why insist that there is any step, logical or psychological, in the transfer of desire from an end that is not an action to the actions one conceives as means? It serves the argument as well that the desired end explains the action only if what are believed by the agent to be means are desired (Davidson 1963, pp. 6-7).

That it is generally otiose to spell out the primary reason does not mean, though, that Davidson's thesis can do without the relevant pair of pro attitude plus belief. The relevant pair is what actually rationalizes action, but we don't always need to mention both in giving the agent's reasons. Often we can make sense of the agent's behaviour without mention of the specific primary reason in question or of one of its components, but for Davidson, that does not mean that, had the agent not had the relevant pro attitude plus belief, her action would have still been rationalized: it would not have been (at least by the mental states in question).

The idea, then, is that the fact that agents themselves would self-attribute primary reasons (or parts thereof) even in automatic cases provides us with an argument for always attributing the required unconscious beliefs. So, even though agents might have been unaware of their beliefs, the fact that they can report them afterwards is a reason to think that the agent had the required unconscious belief.

We have already encountered the first consideration against this argument from self-attribution: just as with introspection, can we accept the agent's own version of events? If we take the agent's own explanation of her behaviour as good evidence, then we have to accept a sort of epistemic authority of the agent over her reasons and actions. And the philosopher, at least as much as the layman, cannot just accept what people say about themselves (Tanney (1995, p. 10) puts this point rather nicely).

The concern here is a methodological one: a theory that need not rely on the agent being both truthful and correct about herself is a methodologically superior theory (Pollard (2005) makes a similar point).

It is not only both philosophical and lay common sense that speak against relying on self-attributions. It is also psychoanalysis, which tells us that agents are often mistaken (in denial) about their reasons, in the way of both being unaware of one's actual reasons, and mistaking other considerations for one's actual reasons. Both these kinds of self-deceptions tell us not to trust self-attributions of reasons.[6]

Another reason for being sceptical about agents’ self-attributions is that, when they come in the shape of answers to questions, they might be influenced by the way in which the question has been asked: so that a Davidsonian question might lead a Davidsonian answer. The agent might provide a primary reason only because the question assumed one; but they might have not actually volunteered one.

[6] Here I shall not discuss the interesting question of whether acting from Freudian beliefs and desires constitutes acting intentionally (under that description, that is). I won't do that simply because that kind of behaviour isn't automatic; because, as we already saw in Section 2, we don't have the right kind of access, 'direct' access, to our behaviour in such cases; so that the sense in which we are unaware of it is different from the sense in which we are unaware of automatic actions.

One more reason not to accept self-attributions as evidence comes from everyday language. People are often accused of, rather than congratulated for, 'rationalizing' their behaviour. The accusation is that they make up their reasons *ad hoc* to make sense of their behaviour. Suppose the meal you've just cooked for your friends turns out tasteless because you have forgotten salt altogether. When the complete absence of salt is pointed out to you, you might rationalize your behaviour to your friends by citing health concerns. What has actually happened is that you forgot. What you are doing, there, is constructing a story that will make sense of your actions and get you off the hook: you are trying to avoid responsibility and look good. Concern for your friends' health would rationalize, in Davidson's sense, your actions if you had actually acted from those considerations. But since you didn't, those considerations do not rationalize your behaviour. What you are actually doing, in making up your story – ‘rationalizing’ in the everyday language sense – is, in short, lying.

Note, also, the similarity between the way in which everyday language understands the practice of rationalizing and my general line of argument against Davidson: both in everyday language and in my argument the respective practices that go under the name of 'rationalization' are accused of constructing, rather than reporting, reasons. And a braver philosopher than I am would claim that to be in itself an argument against Davidson.

The crucial point, indeed, is that constructions are not good enough for Davidson's thesis. His view needs descriptions because it needs actual mental states and actual causes. It has been argued that constructions could be good enough to make sense of an agent's behaviour as rational (see Pollard 2003, p. 424). But as long as those constructions would not point to the agent's psychology (her mental states), then they would not be good enough for Davidson's rationalization, even though they might work on a different conception of rationalization and rationality than Davidson's.

There is another interesting case of construction rather than description in which even Davidson would not want to speak of genuine rationalization. In some cases of reflex behaviour, such as eye blinking, I often discover a threat from the way in which my body reacts to it, rather than the other way around. My eyes might blink, and only afterwards will I realize that there was a fly or that a tree branch was too close for comfort. My eye blinking could, indeed, be rationalized by a desire to protect my eyes and a belief that avoiding the tree branch has the property of protecting my eyes. But if those kinds of movements have to be considered genuine reflexes, then Davidson himself would deny that I have actually acted, even unconsciously, upon such pair of pro attitude plus belief; because otherwise my reaction would be an intentional action rather than a mere reflex.

Indeed, this is a case in which it might be the agent herself who self-attributes the relevant mental states to rationalize eye blinking: “Why did you blink?” “Because of the fly”. But, again, what she would be doing is constructing a story rather than reporting the mental states that actually caused her actions. And in this case Davidson himself would accept this, as far as his view would want to allow for genuine reflex movement.

So the agent’s self-attributions of reasons don’t seem to be good enough evidence for thinking that the agent actually acted from the reasons she attributes to herself. Should we then conclude that self-attributions should be disregarded altogether as a guide for understanding an agent’s reasons? Such a conclusion seems too strong: all I have shown is that, when our only ground for concluding that an agent acted for some reason is that the agent thinks or says that she acted for that reason, then we do not have sufficient grounds for attributing that reason to the agent. Obviously, in normal cases we will accept the agent’s version; but that is just because, in normal cases, we will have other elements which substantiate that version (environment, circumstances, what we know about the agent’s past, habits, and preferences, what we know about human nature, other people’s versions, etc).

3.1.1 Alternative stories

There is an independent reason why self-attributions don't support Davidson's argument. Even if, despite what I have argued so far, we accepted self-attributions as good evidence, not all self-attributions support Davidson's view.

First of all, sometimes people don't know their reasons; they actually don't know why they did something. Dennett (1991) offers a nice example of this:

> I was once importuned to be the first base umpire in a baseball game – a novel duty for me. At the crucial moment in the game (bottom of the ninth, two outs, the tying run on third base), it fell to me to decide the status of the batter running to first. It was a close call, and I found myself emphatically jerking my thumb up – the signal for OUT – while yelling “SAFE!”. In the ensuing tumult I was called

> upon to say what I had meant. I honestly couldn’t say, at least not from any privileged position (Dennett 1991, p. 248).

It might be debatable whether Dennett both intentionally yelled “Safe” and intentionally jerked his thumb up. But it appears obvious that he did at least one of those things intentionally. And he hasn’t got a rationalization for why he did either – or, anyway, he cannot self-attribute a rationalization from the inside, through introspection.

Not only does this kind of case suggest that we are not always able to self-attribute an explanation. It also supports what I have already said about construction rather than description: because we don't know what, if anything, went through our mind, we might make it up.

So this was a case in which the agent was actually unable to self-attribute a rationalization. Other times agents are perfectly able to make self-attributions, but still the kind of self-attributions that they might provide do not support Davidson's view. This is the case for things like “For no reason”, “I didn't think”, 'I just did it”, “I did it automatically”, or “I just wanted to”: many of those replies would apply to what Hursthouse (1991) has called 'arational actions' (a list of which I give in Chapter 5, Section 4.4.1). The point here is not that when agents give these kinds of reports, they must have lacked the relevant mental states. As we just saw, agents’ self-attributions can’t establish as much: agents might have had the mental states they self-attribute, and those mental states might have been the actual causes of their behaviour. But the point is that the fact that agents give these sorts of replies weakens the case for the availability of Davidsonian rationalization, in stopping the appeal to reasons.

What I just said might be taken to be incompatible with what Anscombe had to say about this:

> Now of course a possible answer to the question 'Why?' is one like "I just thought I would" or "It was an impulse" or "For no particular reason" or "It was an idle action - I was just doodling". I do not call an answer of this sort a rejection of the question. The question is not refused application because the answer to it says that there is *no* reason, any more than the question how much money I have in my pocket is refused application by the answer "None" (Anscombe 1957, p. 25).

Anscombe might be read as saying then that those sorts of answers do not stop the appeal to reasons, as I was suggesting. I think that would be a mistaken reading: Anscombe is saying that the question is appropriate, but that the answer is, at least this one time, that there was no reason why I so acted. And this is no problem for Anscombe's account, as long as we read it as saying that everytime we act for a reason, we act intentionally.[7] But it would be a problem for her account if we read it as saying that we act intentionally only when we act for a reason - Davidson's account. Because here we have cases in which there is no apparent reason - as in no reason volunteered by the agent - but still we want to say that the agent acted intentionally.

[7] Here I am not offering an interpretation of Anscombe's account; I am only using it to show why those cases might trouble Davidson's. So it might be that Anscombe's view ought not to be understood as saying that 'everytime we act for a reason, we act intentionally'. Having said that, Kelly and Knobe (unpublished) describe the difference between Davidson's and Anscombe's accounts exactly as I have above.

As we saw in the previous section, obviously the agent's story might be wrong. But, on the assumption that what agents report deserves to be taken at *face value*, then we must also accept that often agents make reports that don't look like primary reasons. So not all self-attributions support Davidson's view.

Here a defender of Davidson's theory might want to object that it isn't true that those sorts of stories don't support Davidson's view. Davidson himself discusses this point in at least two places in *Actions, Reasons, and Causes*. He first says: “We cannot explain why someone did what he did simply by saying the particular action appealed to him; we must indicate what it was about the action that appealed” (Davidson 1963, p. 3). Few pages later, he states: “...it is easy to answer the question, 'Why did you do it?' with, 'For no reason', meaning not that there is no reason but that there is no further reason, no reason that cannot be inferred from the fact that the action was done intentionally; no reason, in other words, besides wanting to do it” (ibid, p. 6).

But here it actually doesn’t matter to my discussion whether Davidson can reconcile this apparently troublesome cases with his thesis. What we were looking for was evidence in favour of his thesis. And those cases do not provide any evidence in favour of Davidson: they only can, at best, be made to fit into his picture. But this latter project, once we have established that those kinds of cases don’t support Davidson, is no concern of us.

So far, then, we have shown why it is problematic to take agents’ self-attribution as evidence for the existence and causal role of unconscious beliefs. And that, even if we could, not all self-attributions would support Davidson's view. But there is one final problem with self-attributions which would be in the way even if we accepted them as evidence, and even if we dismissed all the self-attributions that don't fit Davidson's picture: that agents might very well truthfully and correctly self-attribute reasons, but that does not mean that they are pointing to some actual mental states of theirs as causes of their actions.

The point is two-fold: it means, on the one hand, that self-attributions might lend support to the Humean belief-desire model of motivation (according to which both beliefs and desires are necessary to motivate an agent to act, see Smith 1987 & 1996), but that does not mean that they lend support to Davidson's particular version of it, according to which not only is the belief-desire pair necessary to motivate the agent to act, but actually the belief-desire pair causes the agent's actions. So even if agents did self-attribute reasons, that would not mean that they were self-attributing causes.

This point is pretty simple: on the assumption that a Humean need not accept Davidson's causal thesis, then self-attributions do not support Davidson's causal thesis either. Because they do not explicitly point to causes and, on the above assumption, what they point to – reasons – do not necessarily need to be causes.

Secondly, the fact that agents self-attribute reasons does not mean that they necessarily self-attribute mental states. Indeed, that would be assuming internalism about reasons. According to externalists about reasons for action, such as Collins (1997) and Dancy (2000), reasons are not psychological states of the agent, but facts of the mind-independent world in the light of which agents act.[8]

To this second point it could be objected that, actually, self-attributions of reasons do point to the agent's mind and psychology: agents say things such as 'I thought that x', 'I wanted x', etc. And, since I have here assumed for the sake of argument that self-attributions can be taken as evidence, it is not open to me to reply by pointing to my arguments, in the previous sections, against self-attributions. On the face of it, then, self-attributions seem to support internalists about reasons for action such as Davidson.

I don't think so. What externalists such as Collins and Dancy say is that, when an agent says 'I turned the light on because it was getting darker' (or 'I turned the light on because I thought (or 'believed') that it was getting darker'), the reason is not the agent's belief that it is getting darker, but rather the fact that it is getting darker. They do not necessarily deny that the agent has the relevant belief (or some other cognitive relationship with the fact that 'it is getting darker'); but they argue that the reason is not the belief.

And, if anything, in so far as 'I turned the light on because it was getting darker' is a more common expression than 'I turned the light on because I thought (or 'believed') that it was getting darker', then what agents say appears to be on the side

[8] More on externalism about reasons for action in Chapter 5, Section 4.

of externalists; simply because agents, more often than not, point to the fact, not to the belief. But I won’t defend this latter point; here I only wanted to show that self-attributing reasons doesn’t necessarily mean self-attributing mental states; and that therefore self-attributions don’t support Davidson over externalists.

There is a similar, but independent, consideration which also suggests that self-attributed reasons (or, indeed, intentions) might not imply the self-attribution of mental states. Suppose I automatically flip the switch. Suppose someone asks me: “Did you want to turn off the light?” or “Did you intend to turn off the light?”. And suppose that I would answer positively to both questions. Now, a supporter of Davidson might take that to mean that I am self-attributing a pro attitude (or intention) towards turning off the light. But all that I am saying about myself might actually be that I did not have any attitude or desire not to turn off the light, or that I did not intend not to turn off the light.

Take 'P' to be my pro attitude or intention to turn off the light. The defender of Davidson's thesis would then be proposing that my replying positively to the questions means that I am self-attributing the pro attitude or intention that 'P'. But actually all that I might be saying about myself is that '¬¬P'. 'P' and '¬¬P', here, are not at all equivalent, because in saying about myself that 'P' I attribute a mental state to myself. While in saying about myself that '¬¬P' I attribute no mental state at all. But it is not at all obvious that a layman, in answering the ‘why? Question’, will pick up on such a difference; especially if the question, ‘Did you intend to P?’, suggests one kind of answer, ‘P’, rather than the other, '¬¬P'.[9] Davidson, importantly, needs the relevant mental states because he needs causes: so '¬¬P' won't do for Davidson's theory.

In conclusion, I have shown that we ought not to accept agents' self-attributed stories as good evidence for the truth of a philosophical view. And that, even if we did, not all those self-attributed stories would actually lend support to that philosophical view. And finally that, even if we dismissed all the stories that did not support that view, we wouldn’t have actually found evidence for Davidson's particular view, because we would not have found evidence for two of its most controversial aspects: that reasons are causes, and that reasons are ‘in the mind’.

3.2 Unity

In this section I shall consider another possible argument for the attribution of the required unconscious mental states: that the attribution is necessary in order to give a unified theory of agency; and that therefore we should attribute the required mental states – regardless of the evidence for them or lack thereof – in order to have a unified theory of agency. I will show that this consideration isn't sufficient for the acceptance of Davidson's causal thesis.

It might be suggested that we should always attribute unconscious mental states in automatic cases, notwithstanding the (lack of) evidence, in order to have a unified

[9] Here it could also be argued, as I did in Section 3.1, that the question ‘Did you intend to P?’ leads a positive answer, ‘P’, rather than a negative one, ‘¬P’. That might be true; but here I only need the less controversial claim that the question suggests ‘P’ much more than it does ‘¬¬P’: and this seems so, as long as it is true that the question suggests a positive or negative answer – either P or ¬P – over a double negation.

theory of agency; one that puts forward, as Davidson's does, the same kind of explanation for every kind of action. A unified theory does indeed offer many advantages: firstly, economy and simplicity. And a more economical theory, other things being equal, should be preferred. Similarly, a simpler theory, other things being equal, should be preferred.

But there is a more important advantage offered by a unified theory of agency: the idea that a unified theory does justice to the unity of agency, to the fact that all actions, being actions, share something: whatever it is that marks them all as actions. Indeed, throughout this thesis I have never denied that automatic actions are full-blown actions (*contra*, one might think, Velleman 1992 – see my Introduction), and Davidson's thesis would acknowledge them as such, by offering for them the same kind of explanation provided for non-automatic actions.

So Davidson could acknowledge the unity of agency, and he could acknowledge automatic actions as *proper* actions. He could also acknowledge that automatic actions too call for rational enquiry: automatic actions too are subject, to use Anscombe's expression, to the 'why? question'. And all these similarities, it could be argued, outweigh the differences between automatic and non-automatic actions outlined in Chapter 1: attention, awareness, conscious control, effort, deliberation, thought. Because, by my own admission, those differences may not be enough to make automatic actions *lesser* actions.

All these considerations from the unity of agency I accept. What I deny is that these considerations recommend Davidson's view. Davidson's is a very specific kind of unified theory, because it is a causal one. And the need to always attribute mental states depends exactly on Davidson's commitment to reasons being the causes of actions.

The argument from unity does not, I think, justify accepting Davidson's causal thesis simply because it recommends its most famous alternative just as much. In fact, the theories of the relation between reasons and actions that Davidson sets out to criticize in *Actions, Reasons, and Causes* – Ryle 1949, Anscombe 1957, Hampshire 1959, to cite but a few – provide a unified account of agency too; but one in which the relation between reasons and actions is logical rather than causal. But if the argument from the unity of agency recommends these theories as much as Davidson's, then it cannot be an argument in favour of the acceptance of the crucial difference between Davidson's thesis and the so-called 'logical connection argument' alternative: that there are, in every case, mental states that *cause* and rationalize action.

So unity might be a strong enough consideration to recommend *a* unified theory, but it is not sufficient to recommend *Davidson's* unified theory. But then, if one accepts the considerations from unity, a unified theory like Davidson's will at least be better than a non-unified theory. That might be true: but here I was looking for arguments in favour of Davidson's theory, rather than for a comparison between Davidson's theory and a non-unified theory.

On the other hand, if one accepted – as of Section 1.1 - that Davidson's argument was an inference to the best explanation, unity will be a consideration in favour of Davidson's account of intentional action when it is compared with non-unified views. With this latter point I deal in Chapter 5, Section 4.

3.3 Naturalism

There is another possible argument in favour of Davidson's view: naturalism. Taking a naturalistic approach, according to Blackburn, means to "refuse unexplained appeals to mind or spirit, and unexplained appeals to knowledge of a Platonic order of Forms or Norms; it is above all to refuse any appeal to a supernatural order" (Blackburn 1998, pp. 48-49, found in Pollard 2005, p. 70).

It could be argued that we should embrace Davidson's theory because it promises a naturalistic understanding of the relation between reasons and actions, by claiming that reasons cause actions; and therefore acknowledging both reasons' and actions' place in nature (Davidson (1971, p. 44) takes actions to be a subclass of events: "there is a fairly definite subclass of events which are actions" – and I have no qualms with *this* claim of his).

I take naturalism to be a strong consideration in favour of Davidson's view; but an argument in support of Davidson's causal thesis from naturalism wholly relies, just as with the 'inference to the best explanation', on the premise that Davidson's thesis *alone* promises to give a naturalistic understanding of the relationship between reasons and causes. Indeed, this is just like Davidson's admission that "...failing a satisfactory alternative, the best argument for a scheme like Aristotle's [a causal account] is that it alone promises to give an account for the 'mysterious connection' between reasons and actions" (Davidson 1963, p. 11). Had Davidson written *Actions, Reasons, and Causes* twenty years later, he might have even said that his thesis 'alone promises to give a *naturalistic* account of the mysterious connection between reasons and actions'.

But it's just not true that Davidson's is the only kind of naturalism. McDowell, in *Mind and World*, has suggested an alternative naturalism, one developed around "the notion of *second nature*" (McDowell 1996, p. 84): a person's character acquired through upbringing. McDowell calls the process of acquiring and developing a second nature *Bildung*. It involves "initiation to conceptual capacities" and "responsiveness to rational demands". The general idea is that the characteristics of human rationality and reason (what Sellars (1956) calls the *space of reasons*) are at once natural and familiar.

> This should defuse the fear of supernaturalism. Second nature could not float free of potentialities that belong to a normal human organism. This gives human reason enough of a foothold in the realm of law to satisfy any proper respect for modern natural science (McDowell 1996, p. 84).

Here I am not going to discuss nor defend McDowell's version of naturalism. My point is simply that, since Davidson's is not the only version of naturalism, one of the premises of the argument from naturalism in support of Davidson's causal thesis is false: namely it is not true that Davidson's is the only available version of naturalism. Therefore rejecting Davidson's causal account does not amount to a rejection of naturalism.

4. Arguments against the attribution

I have so far argued that there is no evidence for attributing the required unconscious beliefs in every case; and that unity and naturalism aren't good enough reasons either. In this section I show that there are also arguments *against* always attributing unconscious beliefs: namely, that it is always possible to attribute such unconscious beliefs; and that therefore these unconscious beliefs lack the distinctiveness required by explanation.

This point has already been made by McDowell (1978) for what he calls *consequentially ascribed desires*:

> But the commitment to ascribe such a desire is simply consequential on our taking him to act as he does for the reason we cite; the desire does not function as an independent extra component in a full specification of his reason, hitherto omitted by an understandable ellipsis of the obvious, but strictly necessary in order to show how it is that the reason can motivate him... Of course a desire ascribed in this purely consequential way is not independently intelligible (McDowell 1978, p. 79 and 84[10]).

In our case, then, the attribution of the unconscious belief would be simply a consequence of our taking the agent to have acted for some reason x. The ground for the attribution just being – in the absence of any phenomenological evidence and considering self-attributions not to be good enough evidence – the description under which we take the agent to have acted. So that, if in McDowell's case we take the

[10] Page numbers for McDowell 1978 refer to his 1998 collection, *Mind, Value, and Reality*.

agent to have turned left because she wanted to stop at the supermarket, we consequentially ascribe the desire to go to the supermarket. And in our case, if we take the agent to have flipped the switch to turn off the light, we consequentially ascribe the unconscious belief that flipping the switch has the property of turning off the light.

The problem with this, as McDowell points out, is that the unconscious belief logically depends on what it is meant to explain, namely the action description; and therefore, in McDowell's words, the unconscious belief “is not independently intelligible” (ibid). But then the unconscious belief cannot do its job of explaining the agent’s behaviour under the relevant description, namely 'flipping the switch'. Because then the *explanans* – the unconscious belief – would logically depend on the *explanandum*.

My point is, however, slightly different from McDowell's: the unconscious belief does complete the statement of the agent's reason, differently from McDowell's consequentially attributed desires. But this does not change its consequential nature: it is attributed on the sole basis of the action description that we want to explain; and therefore it depends on that action description, rather than the other way around, which would have been the proper direction of explanation (Pollard (2006, p. 9) makes a similar point).

The consequential attribution of unconscious mental states raises another worry for the explanatory power of these unconscious beliefs that Davidson’s thesis needs: that they leave no room for actions that cannot be rationalized, because it is always possible to rationalize any action by attributing the relevant unconscious belief. In short, unconscious beliefs explain too much because they can rationalize any action, even unintentional (or non-rationalized) ones. But if unconscious beliefs cannot distinguish between rationalizable and non-rationalizable (intentional and unintentional) actions, then they cannot offer a distinctive account of intentional (rationalizable) action.

It is like the case, already discussed in Section 3.1, in which I forget to put salt in the meal I was cooking for my friends. It is always open to me as to others, to attribute unconscious mental states that will rationalize what I have done. I have unintentionally left the salt out of the meal. But my leaving the salt out can be easily made into an intentional action of mine by the attribution of an unconscious primary reason: say, again, a pro attitude towards my friends' health, and a belief that leaving salt out is good for my friends' health.

My point is not that unconscious beliefs *are* always attributed consequentially; it is only that unconscious beliefs *can* always be attributed consequentially. They can be consequentially attributed even when they are not attributed at all, and when they are attributed on other grounds. So I am not denying that there can be other grounds for the attribution of unconscious beliefs to the agent, even in automatic cases. I am only pointing out that consequential attribution is always possible. And the fact that it is possible in every case represents a problem for the explanatory power of those unconscious beliefs: the problem being that those unconscious beliefs, when they are consequentially attributed, depend on what they are supposed to explain, the action description. And even when they are not attributed consequentially, since they can always be consequentially attributed, unconscious beliefs are unable to distinguish between rationalizable/intentional actions and non-rationalizable/unintentional actions.

What we would need, in order for that distinction to be made, are cases in which unconscious beliefs can't be attributed; cases in which it is not possible to attribute the relevant unconscious belief consequentially. So an account of intentional action that depends upon unconscious beliefs does not offer a distinctive account of intentional action.

My argument here is not against unconscious beliefs in general: it only applies to the particular unconscious beliefs required by Davidson. Those that, in Stoutland's (1985) words, must "yield" the action description. Also, here I do not pretend to have found a conclusive argument against Davidson's employment of unconscious beliefs in automatic cases. Had I found that, all my previous arguments would be unnecessary. All I wanted to point out was that unconscious beliefs present a problem: and that in order for them to be utilized in a distinctive account of intentional action, one needs to show that there are cases in which those unconscious beliefs cannot be attributed consequentially. And, importantly, I have shown that this task is different from simply showing that there are other grounds for the attribution: because even when the relevant unconscious belief is attributed on different grounds, it *can* be (or could have been) consequentially attributed.

One then starts to worry that unconscious beliefs must be always attributed, even in cases when the attribution is not warranted, simply because the theory needs them. This worry has already been expressed by Dennett (1991) and Pollard (2006). The latter makes the point more explicitly:

> Davidson's putative states of believing... look like the posits of somebody in the grip of a theory, rather than an independent datum being innocently incorporated into a theory whose correctness is still up for grabs (Pollard 2006, p. 9).

The problem is simple: we assume the correctness of a theory, Davidson's, and we use unconscious beliefs to make the exceptions, automatic actions, fit the theory. If one assumes the validity of one's framework, it is understandable that one tries to make anomalies fit that framework. But, apart from pragmatic considerations, it is obviously unacceptable to do that if one's only argument is one's belief in the correctness of the framework; because that’s the very belief that the anomaly challenges. Otherwise we would never have had a Copernican revolution.

Dennett too thinks that we must be careful in not over-applying one model of action explanation:

> Although we are occasionally conscious of performing elaborate practical reasoning, leading to a conclusion about what, all things considered, we ought to do, followed by a conscious decision to do that very thing, and culminating finally in actually doing it, these are relatively rare experiences. Most of our intentional actions are performed without any such preamble, and a good thing, too, because there wouldn’t be time. The standard trap is to suppose that the relatively rare cases of conscious practical reasoning are a good model

> for the rest, the cases in which our intentional actions emerge from processes into which we have no access (Dennett 1991, p. 252).

This warning can be easily applied to automatic actions: we mustn't apply a model developed around relatively rare cases, non-automatic deliberated actions, to cases where it looks as though there are no preceding mental states.

5. A Davidsonian reply

So in this chapter I have shown that there is no evidence for thinking that in every automatic case agents have the unconscious mental states required by Davidson's view. And I have shown that considerations from the unity of agency, and from naturalism, don't support the attribution of those unconscious mental states either. Finally, I have argued that those unconscious mental states are not explanatory, because they depend on the action descriptions they are supposed to explain, rather than the other way around; and further because they can always be attributed in order to rationalize the agent's behaviour, even in unintentional cases.

Now I want to consider a potential reply to my objections that is still open to supporters of Davidson: they could argue that automatic action descriptions might not be intentional under their *automatic* descriptions, but that they will still be intentional under *other* descriptions. Gorr and Horgan (1980, p. 259), for example, say that it is their intuition that subsidiary performances such as those involved in driving are not intentional under their narrow descriptions, say ‘accelerating’, but only under broader descriptions, say ‘driving’ (more on this in Chapter 4, footnote 11).

Take the following case: I automatically flip the switch. My 'flipping the switch' is rationalized by a pro attitude towards 'reducing my carbon footprint', and an unconscious belief that my action, under description 'flipping the switch', has the property of 'reducing my carbon footprint'. Suppose that the supporter of Davidson's thesis conceded to me that, as I have argued in this chapter, we don't have evidence for thinking that the required unconscious belief will be present in every automatic case. Davidson's supporter could then just say that my action might not be intentional under its automatic description, in this case 'flipping the switch'; but that's not too bad, as long as we can say that it is intentional under other descriptions: for example, 'reducing my carbon footprint'.

So, in general, it is open to a supporter of Davidson to reply that automatic actions will always be intentional under some other description; and that it is therefore not much to concede that Davidson's thesis can't account for their being intentional under automatic descriptions. Davidson's supporter would, in short, bite the bullet – and maybe say that it is only few cases that Davidson's theory can't account for.

But this would be a substantial concession: firstly, because, on Davidson's account, unintentional actions too are intentional under some (other) description. So to say that automatic actions are intentional under other descriptions doesn’t say much in terms of intentionality, given that it leaves open the possibility of automatic actions being unintentional under their automatic description; and so it doesn’t even say that they are intentional rather than unintentional. It says, at best, that they are actions rather than mere movements, because they are intentional under at least one description. But what I have been concerned with throughout is not just that when we do something automatically we are acting; but that when we do something automatically we, normally, act intentionally. When I automatically flip a switch I don't just act; I act intentionally: I intentionally flip the switch.

Secondly, and more importantly, because it is this thesis's driving intuition that automatic actions are intentional: which means intentional under their automatic description; because the other descriptions under which the supporter of Davidson might say that automatic actions are intentional might not even be automatic descriptions. Take flipping the switch: one might be able to rationalize it under 'saving the planet from global warming' (and that might not be automatic because my political stance might be the result of much deliberation), and argue that as long as Davidson can show that 'saving the planet from global warming' is an intentional action, it doesn't matter that Davidson cannot show that 'flipping the switch' is an intentional action. But that's not good enough: the intuition, as the surveys also confirm (see Appendix), is overwhelmingly that 'flipping the switch' is an intentional action too.

So if the supporter of Davidson concedes that Davidson's thesis cannot account for automatic actions as being intentional under their automatic descriptions, then they have conceded that Davidson's thesis cannot account for the intuition with which this thesis is concerned. So by proposing what initially appears to be a draw Davidson's supporter would really be conceding defeat.

Conclusion

In this chapter I have argued that Davidson's account of intentional action, according to which an action is intentional only if it is rationalized by a primary reason – composed of a pro attitude plus a belief – which is its cause, does not always work in the case of automatic actions. It doesn't work, I have argued, because there is no evidence for thinking that in all automatic cases the agent has the required unconscious belief. Since the belief must be unconscious, the evidence can't come from phenomenology; but I have also shown that we cannot accept the agent's own self-attributions as evidence; and that even if we did, not all self-attributions support Davidson's thesis. And that even if we only accepted those self-attributions that do appear to support Davidson's thesis, we find that actually they only support something much broader than Davidson's thesis, because the agent's self-attributions don't support the two most distinctive claims of Davidson's thesis: that reasons are mental states, and that mental states cause action.

I have also shown that there is a crucial argument against the attribution of unconscious mental states: that such attribution is only consequential on the action that the mental state is supposed to explain, and that therefore the *explanans* is not independently intelligible from the *explanandum*. A troublesome consequence of this is, for Davidson, that unconscious mental states can't distinguish between intentional and unintentional actions, because they can be consequentially attributed in every case.

In the next chapter I show that the two most-influential non-reductive views of intentional action in the literature, the Simple View and Bratman's Single Phenomenon View, can't be used to defend the intuition that automatic actions are intentional either.

Chapter 4: Bratman and the Simple View

In the previous chapter I discussed the most influential reductive account of intentional action: that of Davidson.[1] In this chapter I turn to the two most influential non-reductive accounts of intentional action: the so-called *Simple View* and Bratman's *Single Phenomenon View*. I show that neither view can be used to defend the intuition that automatic actions are intentional.

1. The Simple View

According to the Simple View (SV), as formulated by Bratman (1987[2]),

E φ-s intentionally only if E intended to φ.

> For me intentionally to A I must intend to A; my mental states at the time of action must be such that A is among those things I intend. I will call this the *Simple View* (Bratman 1987, p. 112).

So, if automatically flipping the switch is to be an intentional action of mine, then I must have intended to flip the switch.[3] The SV is intuitive, and it provides the

[1] Davidson changed his position on reductionism about intentions: while when he wrote *Actions, Reasons, and Causes* (1963) he was a reductionist about intentions, by 1978, when *Intending* was published, Davidson was no longer a reductionist. Mele and Moser (1994) divide the field between reductionists and non-reductionists about intention in the following way (putting themselves among the non-reductionists). Reductionists: Audi (1973), Beardsley (1978), Davis (1984). Non-reductionists: Harman (1976), Searle (1983), Brand (1984), McCann (1986), Bratman (1987).

[2] A formulation of the SV first appeared in Bratman 1984, from which Bratman derived chapter 8 of his 1987's book *Intention, Plans, and Practical Reason*.

[3] Supporters of the SV include Searle (1983), Adams (1986), McCann (1986, 1991, and 1998), Garcia (1990), and Nadelhoffer (2006).

simplest and most economical explanation of intentional action: why was E's φ-ing intentional? Because E intended to φ.[4]

On the other hand, the SV's denial of the intentional character of non-intended actions has always been seen as a consideration against the view:

> It is thoroughly misleading that the word 'intentional' should be connected with the word 'intention', for an action can be intentional without having any intention in it (Anscombe 1957, §1).

This sort of scepticism towards the SV is motivated by scenarios of the following kind: suppose I turn the radio on with the intention to listen to the news. I know that by turning on the radio I will definitely wake up my flatmate, but I don't intend to wake up my flatmate – all I intend to do is listen to the news. Sure enough, the radio wakes my flatmate up. It doesn't seem right to say that I haven't intentionally woken my flatmate because I didn't intend to wake her. Indeed, I intentionally woke her up even though I did not intend to wake her up – that seems to be the most appropriate way of describing what happened.

[4] For what concerns the intuitiveness of the SV, Knobe (2003, 2005, forthcoming) has recently surveyed people's intuitions on the relationship between intention and intentional action, finding that people are more likely to ascribe intentionality for non-intended deeds that are obviously morally reprehensible than for those that aren't: "they seem considerably more willing to say that a side-effect was brought about intentionally when they regard that side-effect as bad than when they regard it as good" (Knobe 2003, p. 193). This appears to show, then, that people's intuitions on the SV largely depend on their moral intuitions. Generally, though, the fact that people are at all willing to ascribe intentionality for non-intended deeds appears to suggest that the layman does not have a particular intuitive commitment to the SV (see also Nadelhoffer 2006). For more information on so-called 'experimental philosophy' – surveys of intuition such as Knobe's, and the ones I present in the Appendix - see their blog http://experimentalphilosophy.typepad.com: a good source of material on recent developments).

Bratman too thinks that a view of intentional action needs to be able to account for bringing about intentionally unintended consequences such as waking up my flatmate. Indeed, Bratman's Single Phenomenon View, which I discuss in Section 2, can deliver on that point, differently from the SV.

Automatic actions appear to be another consideration against the SV. Even supporters of the SV such as Searle (1983) admit that there are performances which do not require the agent to intend them, such as 'spontaneous' actions – "actions one performs… quite *spontaneously*, without forming, consciously or unconsciously, any prior intention to do those things" (ibid, p. 84, emphasis mine) - and 'subsidiary' actions – "even in cases where I have a prior intention to do some action there will normally be a whole lot of *subsidiary* actions which are not represented in the prior intention but which are nonetheless performed intentionally" (ibid, emphasis mine):

> … suppose I am sitting in a chair reflecting on a philosophical problem, and I suddenly get up and start pacing about the room. My getting up and pacing about are clearly intentional actions, but in order to do them I do not need to form an intention to do them prior to doing them… suppose I have a prior intention to drive to my office, and suppose as I am carrying out this prior intention I shift from second gear to third gear. Now I formed no prior intention to shift from second to third. When I formed my intention to drive to the office I never gave it a thought. Yet my action of shifting gears was intentional (Searle 1983, pp. 84-85).

The reader will recall that I have used the same kinds of examples as Searle's when defining automatic actions in Chapter 1. Bratman himself says something very similar about spontaneous automatic actions:

> Suppose you unexpectedly throw a ball to me and I spontaneously reach up and catch it. On the one hand, it may seem that I catch it intentionally; after all, my behaviour is under my control and is not mere reflex behaviour, as when I blink at the oncoming ball. On the other hand, it may seem that, given how automatic and unreflective my action is, I may well not have any present-directed intention that I am executing in catching the ball (Bratman 1987, p. 126).

Both Searle and Bratman suggest that talking of intentions and intending misrepresents the automatic character of the activities in question; but they come up with different solutions. Searle ends up offering another version of the SV: he distinguishes between *prior intentions* and *intentions in action*, claiming that spontaneous and subsidiary actions do not require prior intentions. Every intentional action, though, including spontaneous and subsidiary actions, requires an intention in action.[5] Bratman offers the Single Phenomenon View (SPV), which I present in Section 2.

Notwithstanding the above worries, if the SV were to be applied to automatic actions, one would have to suppose, just as with Davidson, that one's intention to flip the switch, when one automatically flips the switch, is an unconscious intention. Just like we discussed in Chapter 3, if the agent was aware of her intention to flip the switch, then she would be aware of her action under the description 'flipping the switch', and therefore her action, given the lack of attention and awareness condition on automatic action, could not be automatic under that description. Presumably, then, my arguments against the attribution of unconscious mental events from Chapter 3

[5] I do not discuss Searle at length in this thesis because, as I said, I consider Searle's view (1983) a version of the Simple View. But there is one peculiarity of Searle's view that deserves mention: on Searle every intentional action requires, as we said, an intention in action. But Searle's 'intentions in action' do not cause action, as do intentions for authors such as Bratman. Rather, 'intentions in action' are part of the action: they constitute, together with movement, action. So while it is fair to say, according to Searle (1983, Chapter 3), that intention in action causes movement, it is not the case that intention in action causes action.

would apply to the unconscious intentions of the SV too. But here I don't need to use those arguments because I accept Bratman's argument against the SV, to the discussion of which I now turn.

1.1 The Simple View refuted

According to Bratman, "the Simple View is false" (ibid, p. 115). His refutation rests on the famous videogame counterexample.[6] Bratman supposes that a player is playing identical twin videogames at the same time. The scope of the game is to hit a target on either videogame. The two videogames are connected in such a way that when a target is hit on either one, both videogames finish. Also, they are designed so that it is impossible to hit targets on both videogames simultaneously: "If I hit one of the targets, both games are over. If both targets are about to be hit simultaneously, the machines just shut down and I hit neither target… I know that although I can hit each target, I cannot hit both targets" (Bratman 1987, p. 114).

Bratman supposes that, given that the player will win by hitting a target on either videogame, and given the player's skills, the most effective way to win the game is trying at the same time to hit a target on each videogame; so the player decides to do that. The increased possibility of hitting a target on one of the videogames that comes from having a go at both rationally overwhelms the risk of shutting down the game. Now suppose the player hits a target on videogame 1.

> It seems, again, that I hit target 1 intentionally. So, on the Simple View, I must intend to hit target 1. Given the symmetry of the case,

[6] A version of the counterexample, this time with doors instead of videogames, also appears in Ginet 1990.

> I must also intend to hit target 2. But given the knowledge that I cannot hit both targets, these two intentions fail to be strongly consistent. Having them would involve me in a criticizable form of irrationality. Yet it seems clear that I need be guilty of no such form of irrationality: the strategy of giving each game a try seems perfectly reasonable. If I am guilty of no such irrationality, I do not have both of these intentions. Since my relevant intention in favor of hitting target 1 is the same as that in favor of hitting target 2, I have neither intention. So the Simple View is false (Bratman 1987, pp. 14-15).

The idea is that the player hits target1 (t1) intentionally, but that the player did not intend to hit t1. Because, given the symmetry between the two cases, if the player had intended to hit t1, she would have also intended to hit t2. But if the player had intended to hit both t1 and t2, then her intentions would have been inconsistent, given that the player knows (and therefore believes) that she cannot hit both targets. This latter point rests on so-called rational constraints (or belief requirements) on intention.

According to Bratman's rational constraints on intention, an agent can intend to φ only if the agent does not believe that she will not φ. This is the difference between intentions and desires. While it is perfectly rational to desire to φ even if one believes that one will not φ, it is irrational, according to Bratman, to intend to φ if one believes that one will not φ. For example, it is irrational for Ezio to intend to play in the World Cup if Ezio believes – as he should, given his lack of talent - that he will not play in the World Cup. On the other hand it is perfectly rational for Ezio to desire to play in the World Cup even though Ezio believes that he will not play in the World Cup. Indeed, according to Bratman, if an agent believes that she will not φ, then her attitude towards φ can be, at best, a desire; because irrational intentions just are desires.[7]

So if the player intended to hit both targets, she would be guilty of irrationality on grounds of intention-belief inconsistency, because she knows (and therefore believes) that she cannot hit both targets. But the player is guilty of no such irrationality, so the player does not intend to hit both targets[8]; from which, given the symmetry, it follows that the player does not intend to hit t1; therefore the player intentionally hits t1 without intending to hit t1; therefore the SV is false.[9]

It might be objected that Bratman's counterexample might indeed show that the SV - understood as a necessary condition on intentionality - is false; but that this doesn't necessarily mean that the SV can't be used to argue that automatic actions are intentional. Indeed, if all the Bratman-type counterexamples (and indeed any other potential counterexamples) against the SV that one could develop were cases of non-automatic action, then it might still be that the SV would work for all automatic actions.

[7] In the literature there is, famously, a stronger version of rational constraints (defended by Grice (1971) and Harman (1976, 1986)), against which Bratman argues (Bratman 1987, p. 38). According to those stronger constraints an agent can intend to φ only if she believes that she will φ. But we needn't worry about those stronger constraints, because Bratman's argument only needs his weaker constraints.

[8] This point can also be put without any reference to rationality: the player, given her beliefs, cannot intend to hit both targets; she can only desire to hit both targets. So the player does not intend to hit both targets.

[9] Peter Milne has pointed out to me the difference between claiming that the player's intention to 'hit both targets' is inconsistent with the player's belief that she cannot hit both targets, and claiming that the player's intention to hit t1 and the player's intention to hit t1 are inconsistent with the player's belief that she cannot hit both targets. The latter claim is false. What is true, though, is that the player's intention to hit t1 is, given the player's belief that she cannot hit both targets, inconsistent with the player's intention to hit t2. So Bratman's argument still goes through.

But there is no reason to suppose that counterexamples against the SV depend on using non-automatic actions. Indeed, even if the scenario used by Bratman was a non-automatic action, it doesn't look like the scenario would work against the SV only because the action is not automatic. To show this, it will suffice to suppose that 'hitting t1' is an automatic action; and there is nothing that prevents us from supposing that. It might be that the player is so skilled, or that she is so concentrated, that she doesn't pay attention, or need to pay attention, to which target she is firing at: and that therefore she hits t1 without realizing that she hit t1. The nature of the scenario actually makes that quite likely, since the player is firing at both targets at the same time.

So there is no reason to think that Bratman's counterexample works only for non-automatic cases.

1.2 Trying to rescue the Simple View

In this section I discuss five objections to Bratman’s argument against the SV, showing that none works:

1. giving up on rational constraints on intention
2. overbooking
3. conditional intentions
4. redescribing the action
5. time-specific intentions

1.2.1 Rational constraints

The first objection to Bratman's counterexample consists simply in giving up on rational constraints on intention. Because, as we have seen, Bratman's argument relies on the particular version of rational constraints according to which E can intend to A only if E does not believe that she will not A. McCann (1986, 1991), for example, thinks that "unfortunately, all [rational constraints] are false" (1991, p. 205). So not only are Bratman's weak constraints false, but so too are Grice's strong ones.

> There are a number of examples in which it is rational for agents to try to achieve goals that they believe they will not accomplish, and some of the examples involve mutually incompatible objectives. Moreover… when, unexpectedly, such attempts succeed, the sought-after goals are achieved *intentionally*" (ibid, p. 205).[10]

I find this approach very implausible, because one ends up having to claim, as McCann does above, that achievements that are due to luck will be intentional achievements of the agent. So that I intentionally holed in one just because I tried, even though before end I would have acknowledged that I believed that I was not going to hole-in-one (I would have put it quite strongly: "Don't be silly: it's practically impossible to hole in one from here"). Or that I intentionally won the lottery just because I tried to win the lottery by buying a ticket. I think that to claim that those performances are intentional flies in the face of intuition.

But here I will not engage with McCann's objection because, by McCann's own admission, giving up on rational constraints means giving up on what makes

[10] Adams (1986) objects to Bratman's counterexample on these grounds too.

intentions irreducible to desires: “such constraints are of interest partly for their antireductionist implications, since other motivational states, in particular states of desire, are not similarly encumbered” (ibid, p. 204). But then one’s account of intentional action will be a reductive one, like Davidson’s; and I have already argued against such accounts in Chapter 3. So I am happy to accept that Bratman’s counterexample only applies to non-reductive accounts of intentional action which accept, at least, weak constraints on intention.

1.2.2 Overbooking

Sverdlik (1996) argues that it is sometimes rational to hold inconsistent intentions (as in, two or more intentions that are inconsistent with each other, or an intention that is inconsistent with one’s beliefs). He does so by giving the example of overbooking.

> An airline might rationally overbook a flight, knowing that some passengers will not show up... A rational agent, in such a situation, having certain desires that she wants fulfilled, may rationally form intentions which are such that she believes that they will not and cannot all be fulfilled. Nonetheless she is rational in that her having this set of intentions may be her best strategy for getting what she wants. I will call this strategy the overbooking strategy (ibid, pp. 517-518).

Suppose an airline overbooks a plane with 120 seats by selling 125 tickets. Given their statistics on the number of passengers that usually show up, we may suppose that this is the most rational way for the airline to pursue their goal of filling the plane. In this scenario the airline, according to Sverdlik, cannot be deemed irrational despite holding intentions that are inconsistent with each other, namely the intention to board each passenger despite knowing that only 120 passengers can be boarded.

Before showing what I think is wrong with Sverdlik's objection, I must report that Bratman, in the original statement of his argument, had anticipated an objection on the lines of Sverdlik's; although without dealing, specifically, with the case of overbooking that Sverdlik comes up with. Bratman replies to a potential objection according to which the general rational presumption against inconsistency would be overridden by the case in question, because the agent might have strong pragmatic reasons for intending to hit each target, given that such is the best way to pursue her goal of winning the game: "My response is to reject the contention that I must intend to hit each target in order best to pursue the reward. What I need to do is to try to hit each target. Perhaps I must intend something – to shoot at each target, for example. But it seems that I can best pursue the reward without intending flat out to hit each target, and so without a failure of strong consistency" (Bratman 1987, p. 117).

I think that Sverdlik's mistake is to suppose that, in the overbooking scenario, the airline intends to board each and every passenger. It is only true that the airline does not intend for any passenger to be denied boarding. It doesn't need to intend that because, according to its statistics, some passengers will not show up and therefore everybody who shows up will be boarded. But that does not imply that the airline intends to board each and every passenger to which it has sold a ticket. It only intends to board 120 of them; and so there are five that it does not intend to board. Those five cannot and need not be identified in advance. But they are the five that, statistically, will not show up. So in the case of overbooking the airline does not actually have inconsistent intentions, because it does not intend to board each and every passenger to which it has sold a ticket.

Another way of demonstrating my point is that it would be a mistake to describe the airline's intention as a long conjunction of intending to board p1, p2… p125. There is no one in particular that the airline does not intend to board, but that doesn't mean that the airline intends to board each and every passenger. Indeed, the airline's intention should be described as a long exclusive disjunction, composed of all possible combinations of 120 passengers.

Indeed, the upshot of this is that Sverdlik has actually provided another counterexample against the SV: suppose the airline intentionally boards passenger P. If the airline's intentions can only be described as a long exclusive disjunction composed of all possible combinations of 120 passengers, then we cannot ascribe the intention to board P to the airline, because the long disjunction does not imply that the airline intends to board P, since the disjunction will be true even if the airline does not intend to board P.

Sverdlik's objection, then, can't rescue the SV by arguing that it is sometimes rational to hold inconsistent intentions because Sverdlik does not provide a scenario in which the agent (the airline) holds inconsistent intentions.

1.2.3 Conditional and disjunctive intentions

The third objection consists in redescribing the agent's intentions in such a way that they are not inconsistent. Garcia (1990) has proposed to describe the agent's intentions as a conditional intention to hit t1 should she miss t2, and a conditional intention to hit t2 should she miss t1.

> Bratman is right to think the player cannot rationally have both a *simple* (unconditional) intention to hit target 1 and another *simple* intention to hit target 2… However, denying that she has these simple intentions doesn't require us to deny she has a *conditional* intention to hit target 1 should she miss target 2, along with a similar *conditional* intention to hit target 2 should she miss target 1 (ibid, p. 204).

But Bratman stresses very clearly that the agent is playing at both games simultaneously: it is not as if the player tries to hit a target and then, if that doesn't work, she tries to hit the other one. The player could have chosen that strategy, but Bratman, as we have already seen, is very explicit in saying that the player chooses to have a go at both targets at the same time. So the agent's intentions can at best be described as an intention to hit t1 but not t2 and an intention to hit t2 but not t1. But the conjunction $\{[A \& (\neg B)] \& [B \& (\neg A)]\}$ is always false.

An alternative objection, on similar lines, would be to redescribe the agent's intentions as an exclusive disjunction: the player intends to 'hit t1 or t2', but not 't1 and t2': $[(A \vee B) \& \neg (A\&B)]$. It is not irrational for the player to intend to hit either of the targets but not both; in fact, that is just what the player is attempting to do. But the disjunctive intention $[(t1 \vee t2) \& \neg (t1 \& t2)]$ cannot be reduced to the only intention that can save the SV, t1, because the disjunctive intention $(t1 \vee t2)$ is true even when t1 is false. And therefore the truth of the disjunctive intention does not guarantee the truth of the intention required by the SV, t1. So to say that the player had the intention (t1 v t2) does not actually say that she intended to hit t1, which is what is required by the SV (for the same reasons, as we saw in subsection 1.2.2, Sverdlik's overbooking ends up representing another counterexample against the SV).

1.2.4 Redescribing the action

The fourth objection consists in redescribing the action. One might say that there is no need for such a refined description of what the player does as "hitting t1"; and that, for example, "hitting one of the targets" might be a good enough description of the player's behaviour. So that if the SV can account for *that* description of the player's behaviour, then the SV is safe.

The SV, as we have just seen, can account for this alternative description of the player's behaviour: the intention to hit one of the targets, but not both (the disjunctive intention (t1 v t2)) is consistent with the player's beliefs. So the defender of the SV can in fact say that the player hits one of the targets intentionally with the intention to hit one of the targets.

The problem with this objection is that, for all it says, hitting t1 is still intentional. And, again, for all it says, the agent does not intend to hit t1. So, again, hitting t1 is intentional even though the agent does not intend to hit t1, but only to hit one of the targets. So this objection does not actually challenge Bratman's counterexample to the SV. This objection only moves on to show that, under different descriptions, the SV can still work. But no-one was arguing that the SV never works; only that it is false as long as it argues that an intention to A is a necessary condition for A-ing being intentional, because there is at least a case, Bratman's videogame, on which A is intentional despite the agent not intending to A.

So what this objection would have to do is to actually deny that the player hits t1 intentionally. That way this objection would, rather than attacking Bratman's counterexample, deny the intuition upon which the counterexample depends, namely that hitting t1 is an intentional action of the agent. It could then be proposed that the only thing that the player does intentionally is 'hitting one of the targets', while the player does not intentionally hit t1.[11]

Bratman deals with this objection in his original statement of his counterexample against the SV (1987, pp. 117-118). Bratman gives four reasons why it is implausible to deny that the player hits t1 intentionally:

> First, I want to hit target 1 and so am trying to do so. Second, my attempt to hit target 1 is guided specifically by my perception of that target, and not by my perception of other targets. Relevant adjustments in my behaviour are dependent specifically on my perception of that target. Third, I actually hit target 1 in the way I was trying, and in a way that depends on my relevant skills. Fourth, it is my perception that I have hit target 1, and not merely my

[11] Gorr and Horgan (1980, p. 259 say that their intuition about cases such as the performances involved in driving is that those performances are not intentional under the specific descriptions, say 'braking' (in Bratman's case, 'hitting target 1'), while they are intentional under broader descriptions, say 'driving' (in Bratman's case, 'hitting one of the targets'). Gorr and Horgan don't actually argue for this, and they even admit that "we recognize that in this case… intuitions may differ concerning intentionality. Indeed, our own intuitions are not entirely firm one way or the other" (ibid, p. 259).

> perception that I have hit a target, that terminates my attempt to hit it (ibid, p. 118).

Here Bratman has actually done more than he needed in order to reject the objection in question. He has actually shown not only that it is implausible, given the considerations above, to deny that the player hits target 1 intentionally; but also that to only say that the player intentionally hits one of the targets is not enough, because it misses out on the crucial features of the scenario described above.

Stout (2005, pp. 104-105) has denied that the player hits t1 intentionally by denying that the player is trying to hit t1 - which, as we have just seen, is one of four reasons Bratman offers against denying that the player intentionally hits t1.

> [The player] cannot have been trying to hit the left target either (or the right one for that matter)... If he were really trying to hit it he would not have been going for the right target simultaneously. When he succeeded in hitting the left target there was no residual sense of failure in not hitting the right target. This is because he was not trying to hit the right target (or the left for that matter). What he was trying to do was to hit one of the targets. And his method was not to do this by trying to hit both. His method was to try to hit the left target unless the right one got hit first and to try to hit the right target unless the left one got hit first (Stout 2005, p. 104).

But if you "try to hit the left target unless the right one got hit first" and "try to hit the right target unless the left one got hit first" then, at least until either target is hit - and therefore, on Bratman's scenario, until the end of the game - you are trying to hit both targets. This is because until either target is hit, you are trying to hit the left target, because the right one has not been hit. And, until either target is hit, you are also trying to hit the right target, because the left one has not been hit.

1.2.5 Time-specific intentions

The last objection is put forward by Adams (1986). He suggests we distinguish between the player's simultaneously intending to 'hit t1 and t2', and the player's intending to 'hit t1 and t2 simultaneously' (where 'simultaneously' is, importantly, part of the content of the intention). Only the latter, Adams suggests, is inconsistent. The former, if we specify the content of the agent's intentions so that she "plan[s] to hit each target at slightly different times" (Adams 1986, p. 292), is consistent. So we would have to attribute to the agent, supposedly, an intention to hit target1 at time1, and an intention to hit target2 at time2. And given that the intention to hit target1 at time1 implies, supposedly, the intention to hit target1, then if the two new intentions are consistent, then the agent intentionally hits target1 with the intention to hit target1.

The problem is that the agent only needs to hit one target; and that she knows that. Indeed, as we have seen, a good way of describing what the agent intends to do is by saying that she intends to hit either target but not both, or that she intends to hit one of the targets (see Bratman 1987, p. 117). So to intend to hit both, at whatever time, is still inconsistent with the agent's beliefs. Adams might be right in thinking that to intend to hit each target at a different time is not inconsistent with the agent's belief that, as Bratman says (p. 114), if the two targets are about to be hit simultaneously, the game shuts down. But it is still inconsistent with the agent's belief that she cannot hit both targets.

In conclusion, having assessed and rejected five objections to Bratman's argument against the SV, we can conclude with Bratman that the SV is false. So it cannot be used to defend the intuition that automatic actions are intentional.

2. Bratman's Single Phenomenon View

The view that Bratman proposes in place of the SV is his Single Phenomenon View (SPV), according to which "in acting intentionally there is something I intend to do; but this need not be what I do intentionally" (Bratman 1987, p. 119).[12] The SPV shares with the SV the idea that intention is a necessary element of intentional action. But on the SPV E's φ-ing can be intentional even if E didn't intend to φ, as long as E had some intention ψ, and E's φ-ing was in the *motivational potential* of E's intention to ψ.

We can then formalize the SPV as follows:

E φ-s intentionally only if E intended to ψ and φ-ing is in the motivational potential of ψ.

In order for φ-ing to be in the motivational potential of E's intention to ψ, E does not need to actually intend to φ. E can φ intentionally with only the intention to ψ, as long as φ-ing is in the motivational potential of ψ.

[12] Both the Simple View and Bratman's Single Phenomenon View are versions of the 'Single Phenomenon View', which refers to views that appeal to only one phenomenon, intention (as opposed to views, such as Davidson's, that appeal to both beliefs and pro attitudes). For brevity's sake, here I will refer to Bratman's version of the Single Phenomenon View as Bratman's Single Phenomenon View, or just SPV.

> A is in the motivational potential of my intention to B, given my desires and beliefs, just in case it is possible for me intentionally to A in the course of executing my intention to B. If I actually intend to A, then A will be in the motivational potential of my intention. But we need not suppose that if A is in the motivational potential of an intention of mine, then I intend to A (Bratman 1987, pp. 119-120).

So everything that I intentionally do is in the motivational potential of some intention of mine. But in the motivational potential of my intentions there are also courses of action I do not actually intend. Some action φ is in the motivational potential of my intention to ψ, says Bratman, just in case it is possible for me to intentionally φ "in the course of executing" (p. 120) my intention to ψ.[13]

The difference between the SV and the SPV could be crucial for automatic actions, for two sorts of reasons. Firstly, because even if one accepted that my arguments from Chapter 3 also applied to intentions, the SPV would still not be challenged by those arguments; only the SV would be.

My argument in Chapter 3 was that my automatically flipping the switch is intentional only if I had a pro-attitude towards, say, turning on the light, and an unconscious belief that flipping the switch would turn on the light. But, I have argued, there are not always reasons for attributing the relevant unconscious mental states to the agent. Similarly, on the SV, my automatically flipping the switch is intentional only if I had an unconscious intention to flip the switch. So here one

[13] Mele&Moser (1994) propose, I think, a version of Bratman's SPV. They talk of plans including one's φ-ing: "A person, S, intentionally performs an action, A, at a time, t, only if at t, S has an action plan, P, that includes, or at least can suitably guide, her A-ing" (Mele&Moser 1994, p. 229). If one's φ-ing must be included in one's plan without appealing to the SV (which they don't want to do), then it looks as though they will need to appeal to something like Bratman's motivational potential: φ-ing would then be part of the agent's plan without the agent actually intending to φ.

could argue, symmetrically with Chapter 3, that there are not always reasons for the attribution of such unconscious intentions.

But even if one could successfully manage to argue this latter point – which I will not try to do - that would not constitute a challenge to the SPV. This is because on the SPV all I need, for my automatically flipping the switch to be intentional, is an intention to, say, illuminate the room, or read a book, or whatever. And this intention would not even need to be unconscious, because if the agent was aware of her behaviour under the description 'illuminating the room', that would not imply that she was aware of her behaviour under the description in question, 'flipping the switch' – which could then be automatic. So the SPV appears to be more promising than the SV in accounting for the intentional character of automatic action. One note of caution: the SPV would have to spell out motivational potential without appealing back to mental states such as beliefs, otherwise the arguments from Chapter 3 would apply, at least in so far as those beliefs would 'yield', in Stoutland's words, the action description in question; but appealing to beliefs, as I'll show in the next section, is exactly the direction that Bratman appears to take.

The fact that the SPV doesn't require an unconscious intention whose content makes reference to the action description is also the second reason why the SPV appears to be better suited to automatic actions then the SV. As we have seen in Section 1, even supporters of the SV such as Searle admit that, in automatic cases such as spontaneous and subsidiary actions, the agent doesn't need to priorly intend to A for her A-ing to be intentional. It seems implausible that an agent needs an intention to flip the switch for each automatic switch-flipping of hers.

But the SPV isn't as demanding: as long as there is some intention, and such intention is suitably connected with the action through motivational potential, then the action is intentional. So we don't need to suppose that every automatic switch-flipping is preceded by an unconscious intention to flip the switch. We only need, in every case, some intention: the intention to read a book, say. And the claim that, whenever they automatically A, agents always have some intention seems to be not only more economical, but also much less implausible than the claim that they always have an unconscious intention to A.

An example will clearly show the difference in the way in which SV and SPV deal with automatic actions respectively. Suppose Sarah is driving home from work, as she does everyday. This will involve a lot of automatic performances: many gear-shiftings, many signallings, many looks in the rear-mirror, some fiddling with the radio, and so on. All of those activities are intentional, or so goes this thesis's basic intuition. What the SV would have to say, here, is that Sarah's gear-shifting was intentional only if she intended to shift gear; that her signalling was intentional only if she intended to signal, and so on. And, as we said, on the supposition that these activities are automatic, those intentions would have to be unconscious ones. But the SPV can do without all this: it is enough, on the SPV, to suppose that the agent had an intention to go home, and that all those automatic activities were part of the motivational potential of the agent's intention to go home.

It must be said that Bratman himself admits that the above solution will not apply to every kind of action. It might apply, as above, to the subsidiary kind of automatic action, like all those activities that constitute the umbrella-action ‘driving’. But spontaneous automatic actions such as Searle’s (1983) “suddenly get up and start pacing about the room” are importantly different: they don’t appear to be easily reducible to any over-arching coordinating intention, the way in which subsidiary automatic actions are. So if the SPV might have an advantage over the SV when it comes to subsidiary actions, it doesn’t necessarily have the same advantage over the SV for spontaneous actions. And Bratman recognises this point:

> But matters here are complex, and I am unsure whether such a defence can work for all cases. Perhaps there will remain cases of spontaneous intentional action that fall outside my version of the Single Phenomenon View (Bratman 1987, pp. 126-127).

In what follows I will show that Bratman’s caution is still too optimistic: as it turns out, the SPV doesn’t work for any automatic action.

2.1 Motivational potential

The crucial aspect of the SPV is, then, motivational potential. Bratman’s definition of motivational potential must be distinguished from apparently similar definitions: Bratman does not say that φ-ing is in the motivational potential of my intention to ψ only if I intentionally φ in the course of executing my intention to ψ. Nor does Bratman say that φ-ing is in the motivational potential of my intention to ψ only if I φ in the course of executing my intention to ψ. Neither does Bratman say that φ-ing is in the motivational potential of my intention to ψ only if it is possible for me to φ in the course of executing my intention to ψ. Bratman says, instead, that φ-ing is in the motivational potential of my intention to ψ only if it is possible for me to *intentionally* φ in the course of executing my intention to ψ: “it is possible for me *intentionally* to A” (ibid, pp. 119-120 [emphasis mine]).

This gives rise to a quite obvious circularity. If motivational potential is supposed to give us an account of the intentionality of φ-ing in the absence of an intention to φ, and this account itself relies on φ-ing being intentional, then the account is circular. If, in short, the intentionality of φ-ing is the *analysandum* then the intentionality of φ-ing cannot feature as part of the *analysans*. It must be specified, importantly, that the circularity is not in the definition of motivational potential itself, but only in the account of intentional action that motivational potential is supposed to provide, namely the SPV. So the SPV is circular.

There now seem to be two alternatives for a defender of the SPV: either pretending that Bratman hadn’t said “intentionally to A” (ibid, p. 120), and rather work with the following definition, which avoids the circularity: φ-ing is in the motivational potential of my intention to ψ only if it is possible for me to φ in the course of executing my intention to ψ. Alternatively one could grant to the SPV some intuitive but distinct understanding of what it is to act intentionally; one that can be used to arrive at a proper definition of intentional action. In this latter case, the SPV’s definition of intentional action would possibly avoid the circularity, but it would be, at best, incomplete, because it would depend on an intuitive conception of intentional action. So I choose to go with the former option, and I shall pretend that Bratman had not said “intentionally to A”.[14]

We have seen how potentially useful the SPV is for showing how automatic actions are intentional; and it is also very helpful in showing that non-intended consequences that are still attributable to the agent, such as my intentionally waking up my flatmate with only the intention to listen to the news, are intentional actions.[15] But it has now also emerged how such usefulness depends on the SPV’s account of intentional action being quite broad: for some action to be intentional, on the SPV, it suffices that it was possible for the agent to perform that action in the course of executing an intention of hers. One might think that the account is so inclusive that it doesn’t actually explain why some action is intentional.

Adams (1986), for example, made exactly this point, accusing the SPV of failing to answer the following two questions: “1) in virtue of what is the action intentional under that description?, and 2) why is the action not intentional under other descriptions?” (Adams 1986, p. 294). And Bratman himself is the first to concede that motivational potential isn’t explanatory.

[14] I will not bother with the question of exegesis as to whether the alternative reading of motivational potential that I shall be working with - φ-ing is in the motivational potential of my intention to ψ if it is possible for me to φ in the course of executing my intention to ψ - ought to be attributed to Bratman too, or whether it should be considered a new view.

[15] Here I won’t discuss non-intended intentional consequences, but it is worth mentioning that Bratman thinks that the SPV, differently from the SV, can give us a fair account of the Principle of Double Effect (see Nagel 1986, Foot 1967 and 1985). The idea behind the principle is that some things might be permissible if done as a non-intended consequence of some other goal of ours; but impermissible if intended. The SPV acknowledges the difference between doing something intentionally because one intended it, and doing something intentionally despite not intending it (while the SV doesn’t). And therefore the SPV can distinguish between the agent’s involvement in the two cases, without giving up on the idea that, in both cases, the agent acted intentionally – differently from the SV. Bratman doesn’t deal with the moral side of the question; but it is clear how his SPV could be used to drive a moral wedge between the two cases.

> That my intention includes hitting target 1 in its motivational potential, even though it is not an intention to hit target 1, does not itself *explain* why it is true that I hit target 1 intentionally… The notion of motivational potential is intended to *mark* the fact that my intention to B may issue in my intentionally A-ing, rather than to explain it. It is a *theoretical place-holder*: it allows us to retain theoretical room for a more complex account of the relation between intention and intentional action while leaving unsettled the details of such an account. Such an account would not itself use the notion of motivational potential but would, rather, replace it with detailed specifications of various sufficient conditions for intentional conduct (Bratman 1987, p. 120).

The account that will eventually replace motivational potential is what is required to support any claim about intentionality of action, and therefore, if it is to help us, also my claim about the intentionality of automatic action. Unfortunately, Bratman does not provide such an account. This has already been noted by Mele: "Bratman does not attempt a fully detailed account of intentional A-ing not produced by an intention to A" (Mele 1988, p. 633).

The closest Bratman gets to an actual account of intentional action is an incomplete statement:

> Generalizing, we can expect a full theory of intentional action to generate true statements along the lines of:
> If S intentionally B's in the course of executing his intention to B, and S believes that his B-ing will result in X, and his B-ing does result in X and ________, then S intentionally brings about X.
> For present purposes we can leave aside the subtle issue of just how the blank should be filled in (Bratman 1987, p. 123).

Bratman's overall goal in *Intention, Plans, and Practical Reason* is not to give an account of intentional action, but to provide a theory of intention, his *planning theory* of intention. Therefore an incomplete account of intentional action might have been enough for him. Unfortunately for us, this thesis seeks an account of intentional action. Theoretical place-holders might do for Bratman; but they will certainly not do for me.

There is also another problem with Bratman's incomplete account above: it deploys a belief of the agent: "S believes that his B-ing will result in X" (ibid, p. 123). On this incomplete account, then, an action A is intentional only if the agent has some belief according to which action A will result from an intended action of the agent. With Davidson's account, some action A could only be intentional if the agent believed that A would satisfy a pro attitude of hers. On Davidson's account, if the agent's belief were conscious, then the agent would be aware of her action under description A, and then her action could not be automatic under description A. This will then apply to Bratman's appeal to belief too: if the agent's belief that A will result from an intended action of hers is a conscious belief, then the agent is aware of her action under description A, and then her action under description A cannot be automatic. So, then, Bratman's belief too would have to be unconscious, just like with Davidson. But then the arguments from Chapter 3 would apply.

So, even having granted Bratman a way out of the SPV's circularity, we are faced with an unpleasant trilemma: either we accept that motivational potential, and therefore the SPV as a whole, is just a theoretical place-holder for some view which will actually *account for*, rather than just *mark*, "the fact that my intention to B may issue in my intentionally A-ing" (ibid, p. 120). In that case, the SPV cannot be of any use in defending the intuition that automatic actions are intentional, simply because it is not itself an account of intentional action. Alternatively, we could try to fill in Bratman's incomplete statement. But whatever we fill in that account with, it would still rely on the agent's belief. And that belief, in automatic cases, would have to be unconscious. And therefore the arguments from Chapter 3 would apply. Finally, we could take the SPV at face value; therefore as claiming that its criterion - that it was possible for the agent to A in the course of executing some other intention of hers - is a good enough account of intentional action[16]. But both Bratman and his critics (we have seen Adams and Mele) acknowledge that the SPV, understood as such, doesn't actually explain why action A is intentional.

3. An alternative view

Could we make something of the SPV nonetheless? After all, it did promise to be useful in dealing with automatic actions (and also with unintended consequences). We could, for example, propose to understand motivational potential and the SPV in terms of Goldman's (1970) *by relations*. So that φ-ing is intentional just in case the agent intends to ψ *by* φ-ing: understood broadly such that what comes after 'by' doesn't necessarily need to be instrumental to what comes before 'by'; it could also just be a consequence. This would, indeed, be quite similar to Bratman's 'it is possible to φ in the course of executing an intention to ψ'. The crucial difference would have to be that, on such modified account, the sense in which φ-ing is in the motivational potential of the intention to ψ would not just be that φ-ing is one of the

[16] I said 'good enough' rather than 'sufficient' because Bratman, just like Davidson, does not offer sufficient conditions for intentional action: if they did, as we have already said, then their accounts of intentional action would be subject to the deviant cases objection (more on it in Chapter 5).

indefinite number of things that the agent could do in the course of executing her intention to ψ.

On this modified account, on the other hand, φ-ing would have to be specified in the content of the agent's intention, without being itself intended by the agent. It could for example be proposed that, despite the fact that we don't intend what we do automatically, we nevertheless expect to perform those actions. So that I will not have an intention to flip the switch every time that I flip a switch; but only, say, some general intention: going to bed, or reading a book. But nevertheless I know that things like going to bed or reading a book will involve some switch-flipping. And it might be proposed that this is the sense in which switch-flipping is part of the content of my intention to go to bed without being itself something that I intend: the difference could be spelled out in terms of the kind of attitude that an agent has towards switch-flipping as opposed to, say, going to bed: the agent intends to perform the latter, but only expects to perform the former. This, in turn, could be spelled out in terms of the difference in the attention and thought that the two activities require.

It might be objected that this view, even if it can be made to work, won't be able to account for spontaneous actions. This might very well be true, but then Bratman had already admitted, as we have already seen (1987, pp. 126-127), that his view probably couldn't account for all spontaneous actions either. Bratman's view, we then found, couldn't actually account for any automatic actions. While this modified view might at least account for some.

This modified account would then have some explanatory power, as opposed to the SPV, because it could answer the two questions set out by Adams (1986):

1) In virtue of what is the action intentional under that description? Action A is intentional under description φ-ing because the agent intended to 'ψ by φ-ing' – where the *by relation* is cashed out in terms of the agent's knowledge and expectations over what ψ involves.

2) Why is the action not intentional under other descriptions? This is best answered with an example. Suppose I turn on the light by flipping the switch. This view would say that I intentionally flipped the switch because I intended to turn on the light by flipping the switch. Now, suppose also that I, unbeknownst to me, alert a prowler outside. This view would say that my action was not intentional under the description 'alerting the prowler' because alerting the prowler wasn't something that I expected as a result of turning on the light.

So this modified version of Bratman's view appears to have some merit. The problem is that, as long as it distinctively specifies, as an intention to 'ψ by φ-ing' does, φ-ing as part of the content of the agent's intention, this view too will run against the counterexample that Bratman devises against the SV. This is because it will be irrational for the agent to intend to, say, win the game by hitting t1 and hitting t2, given that the agent knows that she cannot win the game by hitting t1 and hitting t2.

The reason why Bratman's SPV survived the counterexample was that, by only requiring that it was *possible* for the agent to φ in the course of executing an intention to ψ, it did not give rise to an inconsistency. It is perfectly rational, indeed, for both hitting t1 and hitting t2 to be in the motivational potential of the player's intention to win the prize; just because all this says is that it is possible for the agent, in the course of executing her intention to win the prize, to hit t1; and it is possible for her to hit t2. And those two, being mere possibilities, do not give rise to an inconsistency. Yes, it is not possible for the agent to both hit t1 and t2; but that does not mean that, in the course of executing her intention to win the prize, it is not possible for the agent to hit t1; and that it is not possible for the agent to hit t2.

It won't help to point out that this alternative view is using expectations rather than intentions. Because the natural way to understand expectations is in terms of beliefs, and then expectations would be subject to rational constraints too. Because it would be irrational for the agent to expect to hit t1 and to hit t2, when the agent knows that she cannot hit t1 and t2. Here one could propose that automatic actions should not be taken to be part of what the agent actually believes to be involved in the execution of her intention, but only of what it is reasonable for the agent to believe as to what will be involved in the execution of her intention.

This alternative has the advantage of not attributing any belief to the agent – belief that might be subject to the arguments from Chapter 3. But unfortunately speaking of what would be reasonable for the agent to believe still does not get us around Bratman's counterexample: it would indeed be unreasonable for the agent to believe that she will win the game by hitting t1 and hitting t2, because, again, the agent knows that she cannot hit both. This latter claim depends upon the assumption that, even though there might be rational beliefs, intentions, or actions that are, nevertheless, unreasonable (such as drinking a can of paint if one wants to drink a can of paint), irrational intentions, beliefs, or actions will normally be, also, unreasonable ones to hold. So, that, *prima facie*, intending to sunbathe in the Meadows is a perfectly reasonable attitude. But intending to sunbathe in the Meadows on a rainy-Thursday, when one knows that it is raining, is not only irrational, but also unreasonable. I shall not defend this assumption here.

We now see that Bratman, with his counterexample, has refuted the SV, but he has also condemned his own SPV because the only way in which the SPV escapes the counterexample, as shown above, is by giving up on explaining why φ-ing is intentional. Bratman has set the bar so high that even he can't jump it, at least as long as he aspires to deliver an account of intentional action.

The obvious way of lowering the bar is to give up on rational constraints; because it is only due to those constraints on intention that the counterexample arises. But the problem with this strategy, as we have already seen in Section 1.2.1, is that it lands us back with a Davidsonian belief-desire view. If there is no difference, in terms of rational constraints, between intentions and other pro attitudes such as desires, then our account of intentional action will be no different from a belief-desire model, and we have already seen in Chapter 3 what the problem with that sort of account is.

One could propose to give up on rational constraints in a different way. Instead of giving up on rational constraints on intention altogether, one could distinguish between the explicit content of one's intentions, and the implicit content of one's intentions, such that, in my intention to 'ψ by φ-ing', ψ-ing is the explicit content of my intention, while φ-ing is the implicit content of my intention. One could then propose that rational constraints only apply to explicit content, but not to implicit content; so that it would not be irrational for me to intend to 'win the game by hitting t1 and hitting t2' simply because rational constraints do not apply to whatever comes after 'by'. This solution would have the advantage of accounting for the intuition that I do not explicitly intend to perform all the actions, many of those automatic, that are instrumental to – or anyhow part of – the satisfaction of my intentions.

The issue with this proposal is, again, how to cash out the implicit content of one's intentions. If we cash it out in terms of expectations then, as we have seen, we run into the arguments from Chapter 3. If, on the other hand, we cash it out in terms of Bratman's motivational potential, then we give up, by Bratman's own admission, on explaining why φ-ing is intentional. So even though talking of implicit content goes some way towards acknowledging our intuitions, we are still faced with more of the same problems: we still haven't got the account of intentional action we need in order to defend the intuition that automatic actions are intentional.

Conclusion

In this chapter I have looked at the two most influential nonreductive accounts of intentional action: the Simple View, according to which A-ing is intentional only if the agent intended to A; and Bratman’s Single Phenomenon View, according to which A-ing is intentional only if A-ing is in the motivational potential of some intention of the agent. I have shown that neither view works: the Simple View is refuted by Bratman’s identical twin videogames counterexample; which, I have shown, survives five objections to it. Bratman’s Single Phenomenon View, on the other hand, faces a trilemma (even ignoring its evident circularity): either it is just a theoretical place-holder that stands in place of a view to come, in which case it cannot be used to argue for the intentional character of automatic actions. Or we could try to fill in the incomplete view that Bratman presents us with; but, given Bratman’s appeal to belief, any such attempt at filling in the SPV would run against the arguments I have presented in Chapter 3. Finally, we could take the view at face value; but in that case, by both Bratman’s and his critics’ admission, the SPV would not be explanatory. In the last section I have shown that attempts at modifying Bratman’s view fail to provide us with a view which escapes the trilemma. In conclusion, then, we are still at a loss for a view of intentional action that can be used to defend the intuition that automatic actions are intentional. In the next chapter I will develop such a view.

Chapter5: Frankfurt and the 'guidance view'

In the previous two chapters I have argued that Davidson’s and Bratman’s causal accounts of intentional action, and the Simple View, cannot be used to defend the intuition that automatic actions are intentional. In this chapter, developing from Frankfurt’s (1978) work on *guidance*, I will present a way in which the idea that automatic actions are intentional can be defended, the 'guidance view': *E φ-s intentionally iff φ-ing is under E's guidance*. Then in the last section, Section 4, I defend the ‘guidance view’ from four potential objections.

1. Frankfurt

In *The Problem of Action* (1978), Frankfurt famously criticises causal accounts of action, and presents a way in which action can be understood non-causally. This is through the idea of *guidance*. Frankfurt’s is an account that does not rely on the antecedents of actions, and it therefore does not depend on psychological states - intentions (Bratman) or primary reasons (pro attitudes and beliefs - Davidson) – as the causes of action, as the causal theory does. On the other hand, it focuses on the relationship between an agent and her action at the time of action.

> What is not merely pertinent but decisive, indeed, is to consider whether or not the movements as they occur are *under the person’s guidance*. It is this that determines whether he is performing an action. Moreover, the question of whether or not movements occur under a person’s guidance

> is not a matter of their antecedents (Frankfurt 1978, p. 45 – emphasis in the original text).

This is why Frankfurt's view appears *prima facie* a very good one for arguing for the intentional character of automatic actions: because it does not depend on attributing psychological states to agents in automatic cases; attributions that, I have argued in Chapters 3 and 4, are not always warranted.

Frankfurt initially distinguishes between two kinds of purposive movement: purposive movements which are guided by the agent, and purposive movements which are guided by some mechanism that cannot be identified with the agent.

> When we act, our movements are purposive. This is merely another way of saying that their course is guided. Many instances of purposive movement are not, of course, instances of action. The dilatation of the pupils of a person's eyes when the light fades, for example, is a purposive movement; there are mechanisms which guide its course. But the occurrence of this movement does not mark the performance of an action by the person; his pupils dilate, but he does not dilate them. This is because the course of the movement is not under *his* guidance. The guidance in this case is attributable only to the operation of some mechanism with which he cannot be identified (Frankfurt 1978, p. 46).

So not all purposive movement is action because, even though all purposive movement is guided, not all purposive movement is under *the agent's* guidance. For cases of purposive movement that are guided by the agent, Frankfurt proposes to employ the term 'intentional'. "We may say, then, that action is intentional movement" (Frankfurt 1978, p. 46).

Through the idea of purposive movement, Frankfurt gives us an insight into what *the agent's guidance* is:

> Behaviour is purposive when its course is subject to adjustments which compensate for the effects of forces which would otherwise interfere with the course of the behaviour, and when the occurrence of these adjustments is not explainable by what explains the state of affairs that elicits them. The behaviour is in that case under the guidance of an independent causal mechanism, whose readiness to bring about compensatory adjustments tends to ensure that the behaviour is accomplished. The activity of such a mechanism is normally not, of course, guided by us. Rather it *is*, when we are performing an action, our guidance of our behaviour (Frankfurt 1978, pp. 47-48).

For some movement to be under the agent's guidance, then, the adjustments and compensatory interventions don't need to be actualized; it is just a question of the agent being able to make those adjustments and interventions: "whose readiness to bring about compensatory adjustments tends to ensure that the behaviour is accomplished" (ibid.).

This latter point finds confirmation in Frankfurt's famous car scenario, where he stresses that guidance does not require those adjustments and interventions to take place; it only requires that the agent be able to make those:

> A driver whose automobile is coasting downhill in virtue of gravitational forces alone might be satisfied with its speed and direction, and so he might never intervene to adjust its movement in any way. This would not show that the movement of the automobile did not occur under his guidance. What counts is that he was prepared to intervene if necessary, and that he was in a position to do so more or less effectively. Similarly, the causal mechanisms which stand ready to affect the course of a bodily movement may never have occasion to do so; for no negative feedback of the sort that would trigger their compensatory activity might occur.

> The behaviour is purposive not because it results from causes of a certain kind, but because it would be affected by certain causes if the accomplishment of its course were to be jeopardized (Frankfurt 1978, p. 48).

So Frankfurt's view does not depend upon those adjustments and interventions in the same way in which the causal view depends upon psychological states. As Frankfurt explicitly says, those adjustments and interventions might never actually have any causal effect upon some movement, but that does not mean that the movement is not under the agent's guidance; and therefore it does not mean that the movement is not an action. On the other hand, as we have seen extensively in the previous two chapters, it is a necessary condition on the causal view that some movement be caused by particular psychological states, in order for it to be an action.

We have so far individuated two major differences between the causal view (in Davidson's and Bratman's versions) and Frankfurt's view. Frankfurt's view does not depend upon the antecedents of action, as the causal view does: it depends upon the relationship between an agent and her action at the time of action - while she is performing it. Also, Frankfurt's view does not depend on some event, in the form of adjustments and interventions from the agent, actually taking place, as the causal view does – in the form of an intention or primary reason. It depends on the agent's ability and readiness to make those adjustments and interventions.

The reason why the causal view could not be used to claim that automatic actions are intentional was that the causal view depended upon the attributions of psychological states as causes of action. Frankfurt's view does not depend on those: so, in this first respect, Frankfurt's view appears to be one that we could use in order to claim that automatic actions are intentional, as long as we can claim that automatic actions are under the agent's guidance.

The idea that automatic actions are under the agent's guidance is, as we saw in Chapter 1, already in the literature. Pollard (2003) applied guidance to automatic action through the concept of intervention control:

> For we have the capacity to intervene on such behaviours. This is particularly the case for those automatic behaviours which we have learned. Since there was a time when we didn't do such things, it will normally still be possible for us still to refrain from doing them in particular cases (though perhaps not in general). We intervene by doing something else, or nothing at all, either during the behaviour, or by anticipating before we begin it. In this way habitual behaviours contrast with other automatic, repeated behaviours such as reflexes, the digestion, and even some addictions and phobias in which we cannot always intervene, though we may have very good reason to do so. I call this *intervention* control (Pollard 2003, p. 416).

The reader will have already noticed the similarity between guidance and intervention control from the quote; and actually Pollard, in a footnote at the end of the very passage just quoted, writes: "Frankfurt (1978, p. 46-48) describes a similar kind of control in his opposition to 'causal' accounts of action" (Pollard 2003, p. 416). But there are also differences, which I will illustrate later in this section.

Pollard too, then, stresses the fact that agents can intervene, and that it is not the actual intervention that is required, but our "capacity to intervene on such behaviours" (ibid). And Pollard also makes explicit reference to the fact that this capacity for intervention is retained in the automatic case. So, it seems, if when we act automatically we have the ability to intervene to correct (or inhibit) our movements, then those movements will be actions.

Indeed, I have already argued in Chapter 1 for the intuitive idea that we don't lose our capacity for intervention when we are acting automatically: that we downshift automatically does not mean that we cannot stop ourselves from downshifting if the lights, in the distance, turn green. Similarly, we are able to downshift to 2nd rather than 4th, say, if we are suddenly required to slow down dramatically. And, recall Chapter 1, those interventions and corrections don't need themselves to be automatic. It might be, in fact, that the agent intervenes non-automatically, that she pays attention. But this does not mean that the action, before the intervention, and when, as in most cases, the intervention is not necessary, is not automatic.

Here I don't want to rehearse the arguments given in Chapter 1 in favour of the idea that automatic actions are under the agent's guidance or intervention control. I want to argue that guidance is sufficient for the claim that automatic actions are intentional.

In order to do that, I will first clarify the relationship between guidance and intervention control. Those two concepts are, as we have seen, similar; but they are not identical. And the difference is an important one to my claim that guidance is sufficient for intentional action. Indeed, while Frankfurt (1978) claims that guidance is sufficient for intentional movement, Pollard (2003) does not make any claim about intentionality in relation to intervention control; he only says that intervention control is sufficient for responsibility.[1]

One might think, in fact, that intervention control is too broad a concept to use it for intentionality: that an agent could have performed all kinds of interventions says little, one might think, about whether the agent was acting intentionally. Only some potential interventions give us a clue about whether the agent was acting intentionally. Take the car scenario again: that the agent could have intervened to stop the car, or turn on the lights, tells us something about the agent's general control over the vehicle. But it tells us very little about whether the agent was, say, intentionally driving below the 50mph speed limit. What might suggest that the agent was intentionally driving below the 50mph speed limit is only the agent's ability to intervene to reduce her speed (in case her speed was approaching the limit, say).

This particular kind of potential correction says something about the agent's driving speed which the other potential interventions don't say. That is why one might think that not all potential interventions are relevant to intentional action, but only some. And this is why it is important to keep the idea of guidance separate from the idea of intervention:

[1] For more on the relationship between guidance and responsibility see this thesis's Conclusion.

the capacity for intervention is a pre-requisite for guidance, but only the agent's ability to perform some particular corrections and interventions can count as guidance.

We can talk, then, of having guidance as opposed to actually guiding one's behaviour; and of having intervention control as opposed to actually intervening upon one's behaviour. When the agent intervenes, the agent will be guiding her behaviour. But the agent need not intervene in order to have guidance over her behaviour (note the similarity with the everyday language distinction between 'controlling' and 'being in control'). She only needs to have the ability to make some specific interventions.[2]

In order to claim that guidance is sufficient for intentional action, Frankfurt's position must be clarified. In fact, Frankfurt thinks that being under the agent's guidance is sufficient for intentional movement which, he says, is action; but that being under the agent's guidance is not sufficient for intentional action. For intentional action, Frankfurt holds a version of the Simple View: an action is intentional only if the agent intended it.

> Let us employ the term "intentional" for referring to instances of purposive movement in which the guidance is provided by the agent.

[2] Fisher and Ravizza (1998) talk of 'guidance control' as opposed to 'regulative control'. They say that "guidance control of an action involves an agent's freely performing that action" (1998, p. 31). The difference between 'guidance control' and 'regulative control' is exemplified through a Frankfurt-type case: suppose Sally is driving a dual control car. Suppose Sally takes a right turn, and the instructor lets her take such a turn – but if Sally had been about to do anything other than turning right, then the instructor would have operated the dual controls so to make her turn right. Then, Fisher and Ravizza say, Sally has guidance control over the car, because she is guiding her movements which result in the car turning to the right. But she has no regulative control over the car, because she cannot make it go anywhere other than where it actually goes. It is not clear that 'guidance control' resembles Frankfurt's idea of guidance, because it appears that an agent, on Fisher and Ravizza's account, can have guidance control even if she cannot make any relevant corrections and interventions. So here I will refrain from using the term 'guidance control' for the agent's guidance over her movements.

> We may say, then, that action is intentional movement. The notion of intentional movement must not be confused with that of intentional action. The term "intentional action" may be used, or rather mis-used, simply to convey that an action is necessarily a movement whose course is under an agent's guidance. When it is used in this way, the term is pleonastic. In a more appropriate usage, it refers to actions which are undertaken more or less deliberately or self-consciously – that is, to actions which the agent intends to perform. In this sense, actions are not necessarily intentional (Frankfurt 1978, p. 159).

I have already presented and refuted the Simple View in Chapter 4: but the concept of intentional action that Frankfurt uses above is so restrictive that it is even stronger than just the Simple View. It is a kind of Simple View in which intentions are necessarily conscious or deliberate. Clearly, the general refutation of the Simple View already given in Chapter 4 applies also to this even less plausible version.

So what I have to argue for is a stronger claim than Frankfurt's: not only that being under the agent's guidance is sufficient for action, but also that it is sufficient for *intentional* action. Can we find in Frankfurt any evidence for this latter claim?

First, Frankfurt gives us no reason for distinguishing between intentional movement – action – and intentional action. Furthermore, he gives us no reason for thinking that guidance is not sufficient for intentional action, despite its being sufficient for action. Frankfurt merely says that would be "pleonastic"; and that distinguishing would represent a "more appropriate usage" of the term 'intentional action'.

Indeed, there doesn't seem to be anything wrong with thinking that guidance is sufficient for intentional action: if what one is doing is under her guidance - if one does not intervene or correct her action, even though she could - presumably then she is acting intentionally. Indeed, because the agent has control over what she is doing, we tend to think that what she is doing can be attributed to her; because if the deed wasn't to her satisfaction, then she could change it, or prevent it, or stop it, or modify it – she could, in short, intervene, because she is in control.

I think that there are two difficulties with Frankfurt's position: he wants to defend the idea that there is "nothing in the notion of an *intentional movement* which implies that its occurrence must be intended by the agent" (ibid., emphasis mine); I agree with that. But Frankfurt does not say what it is in the notion of an *intentional action* that implies that its occurrence must be intended by the agent. Indeed, it for example cannot be the very term 'intentional', and its relation with 'intention', because both 'intentional movement' and 'intentional action' share that term. And that is exactly why Frankfurt ends up with the tricky, phonetically if not semantically, claim that an intentional movement can be intentional or not intentional.

Indeed, if the issue were just a terminological one, there probably would be a much better term to express the presence of an intention: 'intended'. Those actions – intentional movements – that are under the agent's guidance, but which are not preceded by an intention, would merely be *intentional* actions. Those intentional actions that are not only under the agent's guidance, but are also preceded by an intention, would be *intended* actions. This version would even fit better with Frankfurt’s text, which talks of actions “being intended by the agent” (ibid), rather than being preceded by an intention. And it would not land Frankfurt with the awkward claim that an intentional movement can be intentional or not intentional, but just with the less confusing claim that an intentional movement can be intended or not intended.

The very existence of two different words, 'intentional' and 'intended', suggests that there might be a difference. And Frankfurt, in conceding that 'intentional movement' does not require an intention, acknowledges that we can make sense of something being intentional without necessarily referring to an intention.

But there is a more fundamental problem with Frankfurt’s attempted distinction: he slips back into the same considerations he is criticising, on the very same pages, in attacking the causal view. Frankfurt, as we have already seen, criticises the causal view for assessing a movement not because of itself and its relation to the agent at the time of action, but because of its antecedents. But now Frankfurt has also accepted that intentional actions are constituted by intentional movements: “When a person intends to perform an action, what he intends is that certain movements of his body should occur. When these movements do occur, the person is performing an intentional action” (ibid.). So there is nothing about the movements themselves that makes them intentional actions; what distinguishes them as intentional actions is just that they have been preceded by an intention. But this is exactly the sort of argument that Frankfurt rejects in the case of the causal view.

Clearly, it is open to Frankfurt to adopt the causal view for intentional action while rejecting it for intentional movement – action. But choosing to do so would, in the absence of Frankfurt's motivations for doing so, inevitably cast some doubt on his commitment to his argument against the causal view. More importantly, Frankfurt's attempt to distinguish between intentional movement and intentional action lands him with a version of the Simple View – therefore subject to Bratman's refutation.

This tension should not come as a surprise, given Frankfurt's understanding, with which *The Problem of Action* starts, of what the aim of the philosophy of action should be:

> The problem of action is to explicate the contrast between what an agent does and what merely happens to him, or between the bodily movements that he makes and those that occur without his making them (Frankfurt 1978, p. 42).[3]

So if Frankfurt accepts that the fundamental distinction is the one between action and mere bodily movements, why does he care to draw a further, problematic, distinction, between action and intentional action? Frankfurt has individuated in guidance the element that distinguishes between mere bodily movements, which lack guidance, and action, which is under the agent's guidance. So he has solved, if guidance works, what he considers to be the problem of action. Why go further? In fairness to Frankfurt, he

[3] This is not the only place in his work where Frankfurt makes this point: "events that are actions, in which the higher faculties of human beings come into play, and those movements of a person's body - instances of behaviour other than actions, or mere bodily happenings - that he does not himself make" (Frankfurt 1988, p. 58).

doesn't much care for the further distinction: in an 11-page article, he only dedicates two paragraphs – less than half a page – to drawing the distinction between intentional action and intentional movement.

But that distinction might be important to us: the intuition with which this thesis started is not that automatic actions are intentional movements, but that they are intentional actions. So that is the claim that I must defend. Then again, intuition doesn't probably distinguish between intentional action and intentional movement, so could I just settle for the idea that automatic actions are intentional movements – namely, that they are actions rather than just mere bodily movements? That was the deal proposed by the Davidsonian at the end of Chapter 3: I rejected that deal then, showing how just accounting for automatic actions being actions – intentional under other descriptions – isn't enough.

In fact, Frankfurt offers also another option: I could just accept that the sense in which automatic actions are intentional is pleonastic. So my claim that automatic actions are intentional would, on Frankfurt's account, be vindicated in at least two senses.

But there is one general reason why one cannot be content with just showing that automatic actions are intentional movements and, therefore, actions: that is the very common concept, both in everyday language and in philosophical literature, of unintentional actions. Things we do can be actions of ours despite the fact that we do them unintentionally: those cases would mostly amount to errors and consequences that we did not anticipate.

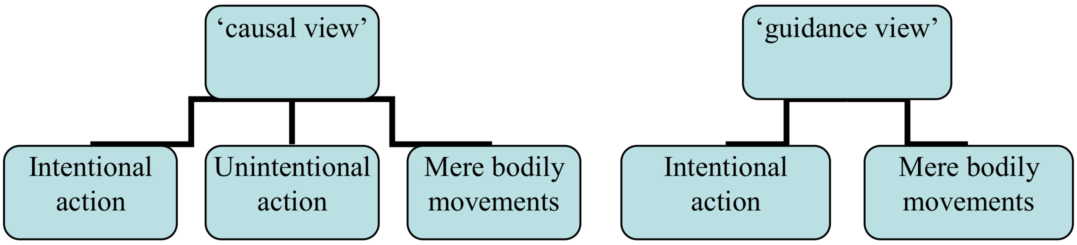
So Frankfurt faces a dilemma – and so do I, as far as I go along with him: proposing his further distinction between intentional movement – action – and intentional action (in a non-pleonastic sense) lands us into the hands of Bratman's refutation of the Simple View, and with the implausible claim that only things we do "deliberately or self-consciously" can be said to be intentional actions of ours – which is no use for the intuition that automatic actions are intentional. On the other hand, renouncing that distinction deprives Frankfurt's view of its capacity to allow for unintentional actions.

In the next section, I will be proposing that we can accept the agent's guidance as sufficient for intentional action without giving up on the distinction between acting intentionally and acting unintentionally.

2. Unintentional actions

A simpler taxonomy of behaviour on the lines of what Frankfurt takes to be a solution to 'the problem of action' does not mean failing to allow for unintentional actions, or so I will argue in this section.

While the causal view presents a tripartite taxonomy, Frankfurt's criterion of guidance, in my modified version, the 'guidance view', only puts forward a bipartite taxonomy of behaviour.



The relevant difference between intentional actions and mere bodily movements is whether the way the agent’s body is moving is under the agent’s guidance - Frankfurt’s criterion. Differently from Frankfurt, though, I propose that guidance is sufficient for intentional action, and that therefore the distinction between action and intentional action does not even get drawn.

Here I shall use Davidson to illustrate the tripartite taxonomy of the causal view: some movement is an intentional action, on Davidson’s (1963) account, if it can be rationalized by a primary reason which is its cause. Now, what further distinguishes in Davidson a movement that is not an action from a movement which is an action (which in turn can be an intentional action or an unintentional action depending on whether it is rationalizable under that description) is whether the movement is intentional under any description. In fact, on Davidson’s account, every action is intentional under at least one description.

What then distinguishes mere bodily movements from unintentional actions is that the latter, but not the former, is intentional under at least one description. To be sure: an unintentional action, for Davidson, will be intentional under a description that is different from the one description under which it is unintentional: “a man is the agent of an act if what he does can be described under an aspect that makes it intentional” (Davidson 1971, p. 46).[4]

One might think that every movement of one’s body can be made to fit into some intentional description; so that for example one could argue that even being pushed was part of my intentionally picking a fight; or my intentionally wanting to be a non-violent individual; or my intentionally leading the ‘good life’. But this would, if anything, be a problem for Davidson’s distinction between unintentional actions and mere bodily movements; but it wouldn’t help me in dealing with unintentional action.

Also, here I share Davidson’s intuition that being pushed is, normally, importantly different from, say, spilling coffee. In the latter case, but not in the former, the accident is, we could say, goal-directed. Alternatively, we could cash out the distinction in terms of passive movement (being pushed) and active movement (spilling coffee).

Here is one of Davidson’s classic examples of unintentional action: “I flip the switch, turn on the light, and illuminate the room. Unbeknownst to me I also alert a prowler to the fact that I am home” (Davidson 1963, p. 686). On Davidson’s account, ‘flipping the switch’ is an intentional action; ‘turning on the light’ is an intentional action;

[4] This whole scheme depends on Davidson’s action individuation understanding (which, as I said in Chapter 1, I accept), according to which different action-descriptions can refer to the same set of movements.

'illuminating the room' is an intentional action; while 'alerting the prowler' is an unintentional action. The first three action-descriptions are intentional because they are rationalized by some primary reason. Something like: I want to illuminate the room (pro attitude), and I believe that turning on the light will illuminate the room, and that flipping the switch will turn on the light (beliefs) – this is a good example, by the way, of how uneconomical Davidson's causal view is. But 'alerting the prowler' is not intentional because it cannot be rationalized by the above primary reason: I didn't have any beliefs about prowlers, Davidson supposes – because I had no idea that there was a prowler outside - to rationalize my alerting the prowler. That is why 'alerting the prowler' was not intentional.

Nevertheless, though, on Davidson's account 'alerting the prowler' is an action because it is intentional under at least one description: indeed, in this case it is intentional under at least three, 'flipping the switch', 'turning the light on', 'illuminating the room'. Because those descriptions belong to the same action as 'alerting the prowler', in that all four descriptions individuate the same set of movements, then 'alerting the prowler' is an action despite not being something I do intentionally.

Some cases that the causal view takes to be unintentional actions aren't, I think, actions at all. Some of them are merely bodily movements. This is because some of the cases that the causal view would consider to be unintentional actions are actually not under the agent's guidance.

Take the following case: suppose I invite you around for coffee. Suppose you are wearing the new pair of trousers that I have just bought you. Suppose I trip over while I am passing you the cup of coffee, and spill the coffee on your new trousers, ruining them. That is the sort of thing of which we want to say that it was not intentional. I really did not want to do it; I was really grateful that you were wearing those trousers; and now I am really sorry. This is a case that Davidson's view would call 'unintentional action': I intentionally offered you coffee, and unintentionally spilled it – so my movements were intentional under the description 'offering coffee' and unintentional under the description 'spilling coffee' – but, under both descriptions, my movements are cases of action: indeed, they are two different descriptions of the same action - of the same set of movements.

It might be that the agent has been careless; or that the agent was trying to do too many things at once; or that she was trying to impress her guest. All those cases are likely to be ones in which the agent could have intervened to prevent the accident, and therefore cases in which the agent's movements were under her guidance. But on the other hand, there could have been an earthquake, or a blackout, or the cup might have been slippery because it hadn't been dried properly; or simply the agent might have been distracted by something that happened at that moment. In those latter cases there is nothing the agent could have done, at the time of action, to avoid spilling coffee: the agent was overwhelmed by nature. In those latter cases, then, her movements were not under her guidance, because she could not have intervened or corrected her behaviour; therefore they were not even actions.

Treating those cases in terms of guidance, rather than in terms of psychological states, helps us to acknowledge, importantly, that some errors are not even actions of ours: we didn't do them, they just happened to us; there wasn't anything we could have done in order to avoid them.[5] So I think that, for the sorts of cases above, my *guidance view* is better than Davidson's causal view.

But I don't think that cases such as 'alerting the prowler' resemble the kind of coffee spilling case in which the agent is overwhelmed by nature: in the coffee spilling case, we say that the agent was overwhelmed by nature because nature took control away from the agent, so that there was nothing that the agent could have done to avoid spilling coffee; so that spilling coffee would, in those particular cases in which the agent is overwhelmed by nature, resemble being pushed. Therefore it would be mere bodily movement, rather than action.

But 'alerting the prowler' doesn't work like that. The movements that constitute 'alerting the prowler' – which are the same movements that constitute the other action descriptions, like 'flipping the switch' – are under the agent's guidance: the agent can correct them, intervene upon them, inhibit them; the agent can do otherwise, or nothing

[5] This kind of idea can be also found in the psychological literature, as for example in Reason's (1990) distinction between *slips* and *mistakes*: I might inadvertently elbow someone while trying to reach Marc to punch him (slip); but I might also punch who I take to be Marc, while the person I punch turns out not to be Marc (mistake).

at all. She is, in short, in control – differently than with being pushed or spilling coffee, as described above.

But I think that there still is a difference in terms of guidance, and one that can distinguish between ‘alerting the prowler’ and ‘flipping the switch’; therefore one that can account for the fact that the agent doesn’t alert the prowler intentionally. The difference is in the kind of interventions and corrections that the agent can make: she can make interventions and corrections such as ‘Let’s not stretch my arm, otherwise I’ll flip the switch’; or ‘Let’s not flip the switch, otherwise the room will be illuminated’; but also things like ‘Let’s not turn the light on, so that I can be a good environmentalist’; or ‘Let’s leave the room in the dark, so that I can save on my electricity bill’. The fact that her bodily movements are under her guidance means that the agent is in control and therefore that, should those or other considerations, but also changes in the environment, occur, she can intervene: correct, redirect, or stop her movements.

All these kinds of considerations can occur while the agent is going through her routine of turning on a light in entering a room. Most times, those kinds of considerations don’t occur, and the agent will automatically turn on the light. But the kinds of interventions and corrections that the agent can make to her performance do not include things such as ‘Let’s illuminate the room to alert the prowler that I am home’ or ‘Let’s leave the room in the dark, so that I can catch the prowler when she tries to come in’. Those kinds of considerations could only apply if the agent knew that there was a prowler, but Davidson is supposing that the agent doesn’t know. Indeed, had the agent known it – and therefore believed it, her 'alerting the prowler' would have, on both my own and Davidson's account, been intentional.

This kind of distinction can be easily applied to Frankfurt's car scenario too. There, it could be said that, say, the agent is intentionally keeping the car below 50mph; because, supposedly, the agent can easily find out what speed she is doing, and intervene upon it if she wants to or needs to. But if we remove the odometer from the scenario, the movements of the agent remain the same; but now, supposedly, the agent is no longer in a position to find out the car's speed, and so we would no longer say that she was intentionally keeping below 50mph.

So there seems to be an obvious epistemic difference between two kinds of cases: in the car case without the odometer, even though the agent's movements are exactly the same than in the car with the odometer, and the agent has guidance over those same movements, the description 'doing 48mph' is not intentional; because, in the absence of the odometer, the agent doesn't know, nor can she know, her exact speed. But that same description is intentional in the case with the odometer, because the agent can check what speed she is doing.

But this epistemic difference does not exhaust all cases: in the prowler scenario, supposedly, the agent can find out that there is a prowler outside; it's just that she would have to go outside and look for it. Admittedly, here there is a difference of degree between finding out your speed by looking at the odometer and finding the prowler by going to the garden to look for her; such that one might want to talk of the agent's ability to *directly* come to know in the case of the odometer, as opposed to the *indirect* way in which the agent would have to go about finding the prowler, by doing other things.

But the crucial point is that the agent has no reason to go outside and look for the prowler, because the agent has no reason to think that there is a prowler outside. Here our epistemic criterion takes on some normative connotations: it looks as though it would be unreasonable to expect of the agent that, every time she was about to flip a switch, she would first go outside to check that no prowler was in the garden. Indeed, we think that people who have these sorts of preoccupations are paranoid (or that they have obsessive-compulsive disorder). The same way in which we would consider paranoid – and dangerous - someone who could not take her eyes off the odometer when driving.

On the other hand, though, it would no longer be unreasonable to expect that the agent go outside in case she hears a sudden loud noise coming from the garden. Similarly, an agent who would react to such noise by going to check the garden would by no means be considered paranoid. This is because the agent, now, would have a reason to go and check. It seems as though rational agents have a background sensitivity to abnormalities, such that they are able to react to them. And, as we have already said in Chapter 1, when the practice has become automatic this capacity to react to abnormalities does not require actively attending to particular aspects of your environment, nor does it require being constantly thinking about potential dangers. The more rational agent appears to be exactly the one who is able to react to abnormalities without having to dedicate all her attentional resources to the circumstances in which those abnormalities could arise but do not arise.

So now we can see the difference between 'alerting the prowler' and 'illuminating the room'. The latter is a description of her movements that the agent can be expected to know or find out at no unreasonable cost – indeed, if the agent ignored that rooms are illuminated by lights being turned on, it would be pretty difficult to make sense of her behaviour. The former, as long as no sudden loud noise comes from the garden, isn't a description of her movements that the agent can be expected to know or find out at no unreasonable cost. If, indeed, a sudden loud noise did come from the garden, and the agent chose to ignore it, then we could argue that the agent alerted the prowler intentionally because she ignored a relevant abnormality.

One can think of this intuitive difference also in terms of which behaviours can be ascribed to the agent, and which can't – for example in terms of praise and blame. It would be reasonable for someone to say to the agent: 'I'm glad you didn't turn the light on. It's good for the environment'; or 'We can't see much, but at least we save on the electricity bill'. But it would be unreasonable for someone to blame the agent: 'I wish you had turned the light on, so we would have scared the prowler away'; or 'You shouldn't have turned the light on, so that we could have caught the prowler'.

We can, therefore, distinguish between the agent's guidance over a set of movements, and the agent's guidance over an action description. The idea is that, with Davidson's 'flipping the switch' scenario, the agent has guidance over the set of movements to which 'flipping the switch', 'turning on the light', 'illuminating the room', and 'alerting the prowler' all refer. On the other hand, while the agent has guidance over action descriptions 'flipping the switch', 'turning on the light', and 'illuminating the room', the agent, for the reasons that I have given, does not have guidance over action description 'alerting the prowler'. So having guidance over a set of movements is necessary for intentional action, but not sufficient: the agent also needs to have guidance over the particular action description. This results in the statement of my 'guidance view' given at the beginning of the chapter: E's φ-ing is intentional iff φ-ing is under E's guidance (where φ-ing is an action description).

If, for clarity's sake, one wanted the distinction between a certain set of movements and the action descriptions that refer to that set of movements to be part of the statement of the view – so that the distinction between guidance over some movement and guidance over some action description was explicit in the view (that's how for example Davidson states his view, see Chapter 3, Section 1), then the 'guidance view' would look as follows:

E's A-ing is intentional, under description φ-ing, iff A-ing, under description φ-ing, is under E's guidance.

In conclusion, I think that appealing to guidance as a sufficient condition for intentional action does not mean that I can't distinguish between when agents act intentionally and when they don't. When an agent cannot be expected to know or find out some description of her movements at no unreasonable cost, because she has no reason to know or find out about that action description, then she cannot be said to be acting intentionally, because she does not have guidance over *that* action description.[6]

3. Deviant causal chains

There is an important difference between my 'guidance view' and the causal view: the causal view only gives necessary conditions for intentional action, while I am proposing that guidance is both necessary and *sufficient* for intentional action. The reason why the causal view stops short of giving sufficient conditions is so-called *deviant cases* (also known as *deviant causal chains*; see Davidson 1973): deviant cases would be counterexamples to the causal view if the causal view were to posit primary reasons (or intentions) as sufficient for intentional action; but deviant cases are no counterexample if the causal view only supposes primary reasons (or intentions) to be merely necessary.

What that means, unfortunately for the causal view, is that it stops short of giving a full account of intentional action. So my proposal does more than the causal view in that it gives both necessary and sufficient conditions for intentionality rather than just necessary ones; also, by posing guidance as a necessary condition for intentional action,

[6] This fits in very well with cases of culpable ignorance and negligence (see, for example, Rosen 2001 and 2003). Those are, indeed, cases in which the agent should have known better, or should have found out before acting. If they have been saying for weeks on the news that the number of prowlers in my area has increased considerably, and I still go on and leave my door unlocked, it is fair to say that I let the prowler in, that I am responsible for it (which obviously doesn't mean that the prowler is any less responsible). 'I didn't know', 'I have forgotten', or 'I never pay much attention to the news' aren't any good as excuses, when you could have been expected to know or find out – at no unreasonable cost - that leaving the door unlocked would let the prowler in. For more on responsibility see my Conclusion.

my proposal can deal with deviant cases: deviant cases, in fact, meet the necessary conditions for the causal view, but not the necessary conditions for my proposal, because such cases are not under the agent's guidance.

The paradigmatic deviant case was set by Davidson in *Freedom to Act* (1973, p. 79) with the climber example: suppose a climber decides to rid herself of the "weight and danger of holding another man on the rope" (Davidson 1973, p. 79).[7] The decision to commit such a horrible act unnerves the climber so much that she loosens her grip on the rope, thereby ridding herself of the other man on the rope. Her decision to rid herself of the other climber both causes and rationalizes the loosening of the rope; but the agent did not intentionally loosen the rope: it was an accident. Such cases would be counterexamples to a view according to which a primary reason causing an action which it rationalizes would be sufficient for that action to be intentional. Nevertheless they are no counterexample to the causal view as long as the causal view does not set sufficient conditions, but only necessary ones.[8]

What is problematic for the causal view, with regards to deviant causal chains, is that the primary reason (or the intention) causes the agent to act in the way she had reasons for (or an intention to) exactly by making her lose control over her action. And that is where

[7] The other famous deviant scenario is the one in which Fred runs over his uncle by accident on his way to kill his uncle (Chisholm 1966).

[8] In footnote 5 of the version of *Actions, Reasons, and Causes* reprinted in Davidson 1980, Davidson says explicitly that he does not want to pose sufficient conditions: "I say 'as the basic move' to cancel any suggestion that C1 and C2 are jointly *sufficient* to define the relation of reasons to the actions they explain. For discussion of this point, see the Introduction and Essay 4 [Freedom to Act]" (Davidson 1963, p. 12). C1 and C2 are, respectively, the necessary condition for primary reasons that I quote on this page (see above), and the claim that "A primary reason for an action is its cause" (Davidson 1963, p. 12).

the counterexample arises: because it looks as though the agent did not have the right kind of control over what happened to say that she acted intentionally, even though what happened was caused by the relevant primary reason (or intention).

That the causalists are troubled by deviant cases shows, interestingly, that some sort of control condition must be implicitly necessary even in their accounts of intentional action, because it is exactly the intuitive lack of control that makes deviant cases troublesome.

On the one hand Davidson wants to say that the agent did not loosen her grip intentionally; because it was an accident: it was something that happened as the result of the agent losing control, rather than something that happened under the agent's control. On the other hand, though, the case matches Davidson's necessary conditions for intentionality: her loosening her grip is caused by her desire to get rid of the other climber, and her belief that by loosening her grip she would get rid of the other climber.

So the causal view, if primary reasons (or intentions) were posed to be sufficient, would be in trouble because the deviant case would be a counterexample in which the event is caused by the relevant primary reason (or intention), but it is, nevertheless, not intentional. That is why Davidson falls short of setting sufficient conditions for intentional action. Davidson admits this very candidly: “"[w]hat I despair of spelling out is the way in which attitudes must cause deeds if they are to rationalize the action" (Davidson 1973, p. 79). And this problem should come as no surprise, if one recalls

Frankfurt's general criticism of the causal view: that it focused on the antecedents of action rather than on action itself.

Proposing guidance as a necessary and sufficient condition for intentional action, on the other hand, does not share the causal view's problem with deviant cases: because it is characteristic of deviant cases that the agent's movements, as in losing grip on the rope as a result of nervous tension, are not under the agent's control: the agent is not guiding her movement; she grows so nervous – because of her evil temptations – that she loses her capacity for intervention; she loses guidance of her movements. Indeed, losing grip cannot even be said to be something the agent does. It is rather something that happens to her, a mere movement of her body, rather than an action of hers. In fact, it belongs to those cases, already discussed, in which the agent is overwhelmed by nature.

Other causalists have attempted to deal with deviant causal chains by including a guidance-type requirement. Here is, for example, Mele&Moser's (1994) proposal[9]: "on our view, the proximal intentions to A whose acquisition initiates intentional A-ings *sustain* and *guide* the A-ings" (Mele&Moser 1994, p. 236). Those authors appear to accept that what is needed to deal with deviance is to supplement the causal connection with a guiding role for the intention. The intention, then, would cause the action not just in the sense of initiating it; but also in the sense of guiding and sustaining it. This is very clear in Thalberg's version:

[9] For other examples of this kind of approach, see Brand (1984), Thalberg (1984), Alston (1986), Mele&Adams (1989), Mele (1992).

> The person's intention only inaugurates a sequence of causally related goings-on which terminate in behaviour; it does not continue to shape events, particularly the behaviour. I think a full-blown causal theory prescribes a tighter hookup – what I call 'ongoing', 'continuous' or 'sustained' causation (Thalberg 1984, p. 257).

These kinds of proposed solutions are, though, subject to a counterexample to which my version isn't subject. The counterexample is sketched by Moya (1990, found in Stout (1996, pp. 86-95)) on the lines of Chisholm's (1966) deviant case. Fred intends to drive over his uncle. He is afraid to miss him, so, on his way there, not to waste any time, he drives over a pedestrian that had got in the way. The pedestrian, sure enough, is his uncle. Here, it looks as though Fred's behaviour is sustained by his intention. There is no gap, as in the original deviant case, which interrupts the guiding or sustaining role of Fred's intention. But still, it looks as though Fred does not run over his uncle intentionally.

But my view does not have a problem with this case: it seems as though Fred has guidance over running over the pedestrian; but he does not have guidance over running over his uncle, because Fred could not be reasonably expected to know or find out, at no unreasonable cost, that the pedestrian was his uncle – simply because Fred had no reason to think that the pedestrian may have been his uncle. Indeed, had Fred been aware that his uncle, at that time, would have been walking down that road, then we might be happy to concede that Fred ran over his uncle intentionally.

Finally, I want to show that my 'guidance view' can also deal with cases of so-called *consequential* deviance (as opposed to the *basic* deviance of cases such as Davidson's climber, see Bishop 1989 and Schlosser 2007). The paradigmatic case of consequential deviance goes as follows: suppose I intend to kill Sam, and so I shoot Sam in order to kill him, but I miss. Nevertheless, the noise from my shot awakens a herd of wild pigs, which trample Sam to death.

The satisfaction of my intention to kill Sam was caused by my intention to kill him. Nevertheless, I don't intentionally kill Sam (indeed, it's not even clear that it's me who kills him; the pigs do). This is another case that satisfies the necessary conditions of the causal view, but that is not, intuitively, an intentional action.

My 'guidance view' can deal with it pretty easily: at the time in which Sam is killed by the pigs, I have no guidance over Sam's killing. I can't directly intervene to stop the pigs. Furthermore, at the time when I fire the gun, I can't be reasonably expected to know or find out, at no unreasonable cost, that I will awake a herd of wild pigs that will trample Sam to death. I therefore don't have guidance over Sam's killing, and that is why Sam's killing is not an intentional action of mine.

4. Objections

In the rest of the chapter I will deal with four objections to the idea that guidance is necessary and sufficient for intentional action.

4.1 Regress

My proposal appears to be subject to a charge of regress; such charge had already been noticed by Frankfurt himself:

> Our guidance of our movements, while we are acting, does not similarly require that we perform various actions. We are not at the controls of our bodies in the way a driver is at the controls of his automobile. Otherwise action could not be conceived, upon pain of generating an infinite regress, as the matter of occurrence of movements which are under an agent's guidance. The fact that our movements when we are acting are purposive is not the effect of something we do. It is a characteristic of the operation at that time of the systems we are (Frankfurt 1978, p. 160).

As Ruben (2003, p. 112) notices, Frankfurt here does not give us an alternative account of guidance that avoids the regress – Frankfurt is merely stating that such an alternative must exist, otherwise guidance cannot be constructed free of regress. That, however, won't do as a reply to the charge of regress – so below is my reply to the charge, in which I argue that my view does not generate a distinctive regress: as in, one to which the opposing proposal, the causal view, isn't also subject.

The charge appears to be that, on my proposal, whether an action is intentional depends upon whether it is under the agent's guidance. And that whether an action is under the agent's guidance depends, in turn, upon whether the agent, at the time of action, is able to intervene over her behaviour, make corrections to her movements, or redirect them. Those activities, in turn, appear to be intentional actions. Therefore, or so goes the objection, my account of intentional action depends itself upon intentional actions.

This objection depends, then, on the claim that guidance is the agent's ability to perform some intentional action.[10] But I think that, if this is the objection, then this objection can be made against the causal view as well, in so far as, on the causal view, intentional action depends on intentional states (psychological states): because proponents of the causal view accept that an essential feature of intentional states is that they are dispositional: not only behaviourists like Ryle (1949), or Stout (1996), but also Davidson: "Primary reasons consist of attitudes or beliefs, which are states or dispositions" (Davidson 1963, p. 12).

The natural interpretation of this passage from Davidson seems to be that "states or dispositions" is an inclusive disjunctive, such that some beliefs will be states, and some dispositions; similarly, some attitudes will be states, and some dispositions. So if my view can be charged with regress for appealing to abilities or dispositions, then Davidson's view can be charged with just the same regress.

[10] And in this respect this *regress* objection is importantly different from the *regress* one that Ryle (1949, p. 67) moves against volitionism. The objection against my account says that on my account whether an action is intentional depends on whether the agent is able, at the time of action, to perform other intentional actions. Ryle's objection is that on volitionist accounts whether an action is intentional depends on whether it was brought about by an act of will – itself an intentional action. So my account of intentional action would depend on the agent's ability to act intentionally; while volitionist accounts would depend on the agent actually acting intentionally through an act of will. On this point, see Stout (2005, p. 9).

One might concede this, but press me on the fact that some other theory that appeals only to states cannot be charged with such a regress – and that therefore my proposal is at least worse off than a theory that only appeals to states.

It seems to me that, if my theory is charged with the regress, even a theory that only appeals to states can be easily charged with that regress: in so far as that theory takes these states to have dispositional properties. Take functionalism: if a state is defined and individuated by its functional properties rather than its intrinsic ones, then some, if not all of those properties, will be dispositional ones. Such and such state tends to cause x; such and such state tends to be caused by y – where x and y can be either mental states or behavioural patterns – intentional actions, for example.

Just to show that I am not re-writing functionalism to fit my purposes, here is a very authoritative account of functionalism, straight out of the '70s, signed by none other than Jerry Fodor and Ned Block:

> But FSIT [Functional State Identity Theory] allows us to distinguish between psychological states not only in terms of their behavioural consequences but also in terms of the character of their interconnections. This is because the criterion of identity for machine table states acknowledges their relations to one another as well as their relations to inputs and outputs (Fodor and Block, 1972, p. 167).[11]

[11] The following quote makes the relation between functionalism and dispositional properties even more explicit: "According to a prominent form of functionalism, the functional state identity theory, mental properties are higher-level, dispositional properties. To be in pain, for instance, is to be in some state or other apt to be caused by bodily damage and apt to cause avoidance behaviour (an actual functional

Those outputs are actions. Block again: "Functionalism says that mental states are constituted by their causal relations to one another and to sensory inputs and behavioural outputs" (Block 1980, p. 1). Just to make sure: it would not be of any help to the supporter of the distinctive regress to distinguish between behavioural outputs and intentional actions, because I am only measuring my view against accounts of intentional actions. So versions of functionalism that deny that intentional actions are one kind of output are not relevant to the charge of distinctive regress.

If one were to develop an account of psychological states that was free of dispositional properties, and then use it for an account of intentional action, and then criticise my account on grounds of regress, then I might have to answer some more questions. Till then, I can reject the charge of regress, because even the ability to intervene is not enough to generate a distinctive regress. One might now ask me to show that the actual intervention does not generate a regress: I won't do that – because my account does not rely on the actual intervention. It only relies on the ability to intervene.

4.2 Principle of Alternate Possibilities

Ironically enough, a potential threat to my proposal comes from Frankfurt himself, with his famous counterexample to the Principle of Alternate Possibilities (PAP). Frankfurt's counterexample is supposed to show that it is a mistake to hold, as PAP does, that an

characterization of pain would be rather more complicated than this). *Mental properties, on this view, are purely dispositional*" (Heil and Robb, 2003, p. 182 – my emphasis).

agent is responsible for an action A only if she could have done otherwise than action A. Frankfurt makes this point by showing cases where, even though the agent has chosen to act in the way she did, the agent could not have acted differently.

One could apply this famous counterexample to my proposal as well. Someone could say that, if some intuitively intentional actions were such that the agent could not have done differently, like in *Frankfurt-type cases* (terminology from Fischer and Ravizza 1998), then my view of intentional action would be false, because there would be actions that are intentional even though they aren't under the agent's guidance. In this section I will show that *Frankfurt-type cases* are no counterexample to my view.

This is how Frankfurt sets out his counterexample against PAP:

> Suppose someone - Black, let us say – wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide to do something other than what he wants him to do. If it does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones decides to do, and that he does do, what he wants him to do. Whatever Jones's initial preferences and inclinations, then, Black will have his way…
> Now suppose that Black never has to show his hand because Jones, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. In that case, it seems clear, Jones will bear precisely the same moral responsibility for what he does as he would have borne if Black had not been ready to take steps to ensure that he do it. It would be quite unreasonable to excuse Jones for his action, or to withhold the praise to which it would normally entitle him, on the basis of the fact that he could not have done otherwise. This fact played no role at all in leading him to act as he did. Indeed, everything happened

> just as it would have happened without Black's presence in the situation and without his readiness to intrude into it (Frankfurt 1969, pp. 835-36).

The agent would have performed the action in question either way; so it is not true that the agent could have acted otherwise. According to PAP, then, the agent is not responsible for doing what she did. But, intuitively, the agent decided of her own will to do the action in question; and it seems that he must be held responsible for the action in question. Therefore PAP is false.

There is, in the literature, an argument against Frankfurt-type counterexamples that I find decisive: it was put forward by Peter van Inwagen in *An Essay on Free Will* (1983). Van Inwagen's counterargument is quite simple (1983, p. 170): in the alternative scenario, where Black intervenes, Black's intervention is in the causal history of what Jones brings about, while in the actual scenario it isn't. If we think that this difference in causal history is sufficient for a difference between the two kinds of action-events, then the action-event in the actual story is different from the one in the alternative scenario. Therefore, van Inwagen says, what Jones brings about in the actual scenario is different from what Jones would have brought about in the alternative scenario. But if that is true, then the alternative scenario does not show that Jones could not have avoided bringing about what he brings about in the actual scenario – because in the alternative scenario Jones brings about something different (or maybe nothing at all, if it's Black who acts – I deal with this point at the end of the section).

The controversial aspect of this counterargument is quite clear: it is the idea that the difference in the causal history of the alternative scenario is sufficient for saying that, in the alternative scenario, Jones would have done something different; and that therefore it is not true that Jones did not have alternative possibilities open to him: the alternative scenario is a genuine alternative possibility, so that PAP isn't, after all, refuted by Frankfurt-type cases.

This aspect of van Inwagen's argument is picked at by Fischer (1994): according to him, the difference is not *robust* enough to ground attributions of responsibility (p. 142).

> … my basic worry is that this alternative possibility is not sufficiently robust to ground the relevant attributions of moral responsibility… it needs to be shown that these alternative possibilities ground our attributions of moral responsibility. And this is what I find puzzling and implausible (Fischer 1994, p. 140).

Even though Fischer says himself "I do not have a decisive argument against it [van Inwagen's strategy]" (1994, p. 140), I think his worries concerning the robustness of the alternative possibility should be met and can be met.

There are at least two things that Fischer could mean by *robust*: Fischer might want an alternative possibility to be robust in the sense that it is different enough from the actual possibility to ground attributions of responsibility. Alternatively Fischer's cry for robustness might be a cry for actuality. His complaint would then be that the alternative possibility is not robust enough simply because it is not actual – and it is on actualities that we should ground attributions of moral responsibility. This latter interpretation can be traced back to Fischer's general compatibilist project. Here I don't want to enter the free will debate more than I need to, nor take a particular stand on it. It is only those who have some commitment to compatibilism that need this restriction to actualities. Since I don't have such a commitment, I will concede to Fischer that, if he restricts robustness to actuality, not even the alternative possibilities described below will convince him.

Back to the first interpretation of robustness: the idea that the alternative possibility isn't different enough from the actual scenario to ground attributions of responsibility. After all, supposing that, for example, what Jones brings about is Fred's death, van Inwagen is not denying that both in the actual scenario and the alternative scenario Fred dies. Van Inwagen is only saying that the event of Fred's death in the actual scenario is different from the event of Fred's death in the alternative scenario. To which Fischer replies that, given that Fred dies anyhow, the difference between those two events is not *robust* enough.

But I think that there is something very robust that van Inwagen could say in replying to Fischer: pretty simply, that, given the different causal histories, the difference is as robust as it can possibly be, because in the actual scenario the event of Fred's death is an action (of Jones); while in the alternative scenario the event of Fred's death is not an action (or if it is, it is not Jones's, but Black's). If this is true, then it is true of Jones in the alternative scenario that he does not kill Fred; while in the actual scenario he does. And therefore Jones, when he kills Fred in the actual scenario, had the alternative of not killing Fred, which is what happens in the alternative scenario. And this is the ground for holding Jones responsible: that he could have done otherwise. Yes, Fred still dies in the alternative scenario – but it is not Jones that kills him, but Black (if anybody). I cannot think of a more robust difference than the one between Jones's killing Fred and Jones's not killing Fred.

If Fischer insisted that the outside intervention from Black is not robust enough to say that Jones did not do it, then he would suddenly be in the implausible position of having to claim that agents act even when they are puppets under the complete control of someone (or something) else. The implausibility of this appears even stronger if one uses Fischer's own version of Frankfurt-type cases, in which "Jack has secretly installed a device in Sam's brain which allows him to monitor all of Sam's brain activity and to intervene in it, if he desires" (Fischer and Ravizza (1998), p. 29). If this doesn't count as Jack's controlling Sam, and making him do things, rather than Sam acting, then I don't know what does. Indeed, on my account it is Jack and Black who act, because they can both intervene upon Jack's movements.

Importantly, here I am not just saying that in the alternative scenario Jones's movements do not constitute, on my proposal, an action of his (in fact they don't because Jones isn't guiding his behaviour, Black is – and therefore it is Black who acts). Here I am also saying that on any view that wants to distinguish between mere movements and actions - and the causal view wants to do that - Jones's movements in the alternative scenario cannot be Jones's actions because they are controlled by Black – and therefore they are, on my view, Black's actions.

It looks as though Black not only acts in the alternative scenario - because he intervenes over Jones's behaviour – but also that Black acts in the actual scenario, given that he has the ability to intervene; and that guidance is sufficient for intentional action. This might seem counterintuitive: Jones, in the actual scenario, acts of his own will – and Black, in the actual scenario, does not do anything: he has the power to intervene, but he does not use it. But none the less, Black's will is being executed through Jones. Also, one might want to suppose, the fact that Black can intervene depends on something Black has done in the past (installing the device; instructing someone to install the device, etc.).

But accepting that Black is acting in the actual scenario poses the problem of Jones's agency: if it is Black who acts through Jones, then Jones is not acting, it would seem. But how can that be, given that Jones is acting of *his* own will? Here I want to propose that in the actual scenario Jones and Black are *both* acting, but that Black is acting *more* than Jones.

The idea is that guidance is not mutually exclusive: so even if Black is guiding x, Jones can be guiding x too, because they can both intervene and make corrections on x. The amount of guidance they will have depends on the number and kind of possibilities for intervention open to them: Jones has very little guidance, because, given Black's device, there is only one possibility open to him: either Jones φ-s or he does not φ. And if Jones does not φ, then the device will intervene. Black, on the other hand, having installed the device, can do all sorts of things.[12]

In the actual scenario, Jones kills of his own will. It was open to him not to kill – and then Black would have intervened. So while it was up to Jones whether he killed or not, it was not up to him whether the victim died or not. While, in Black's case, it was both up to him whether Jones killed or not, and whether the victim died or not (again, this depends on how powerful is the device). That's why Black's got more guidance than Jones, and that's why Black is acting more than Jones.

Simester (1996), someone who accepts the core of Frankfurt's argument for guidance and against causalism, thinks that the kinds of cases above call for a refinement to Frankfurt's guidance.

> …it is not a sufficient condition of behaviour being action that such behaviour is guided, in Frankfurt's terms, by an agent. This is because behaviour can be guided by more than one mechanism. Suppose that Alice takes Bill's hand and smites Chloe on the head with it. Suppose further that Bill is capable of resisting Alice's use of his hand in this way, but refrains from doing so; indeed, Bill is prepared to hit Chloe himself, were Alice not doing it for him. Then on Frankfurt's analysis, both Alice and Bill smite Chloe (with Bill's hand). But this seems wrong. It is Alice who guides Bill's hand: Bill merely allows his body to be acted upon. Bill's deed is one of not resisting and the movement of his hand is (vis-à-vis Bill) a consequential event (Simester 1996, p. 170).

[12] Admittedly, how much guidance Black has does depend on whether the device has been set to redirect Jones's behaviour only in case he is not about to φ or whether the device can change Jones's behaviour in any kind of circumstance.

So Simester proposes to strengthen Frankfurt’s conditions in the following way: “that the behaviour is not caused by guiding forces external to the agent who exhibits it” (ibid). Simester’s intuition seems to be that it is just plain wrong to say that Bill is acting; in a similar way, one might want to say that, when Jones’s behaviour is under Black’s guidance, Jones cannot be said to be acting. Simester admits himself that, on Frankfurt’s account, the natural interpretation would be the one I have given for the case of Black and Jones, such that they are both acting.[13] And I don’t see any reasons why we should reject that kind of reading – apart from Simester’s intuition. Simester himself says that “The fact that Bill's behaviour constitutes an omission might not, in such a case, prevent his being held legally responsible for his behaviour, qua consequence” (ibid). What better way of holding Bill legally responsible, then to suppose that Bill is (partially) acting?

4.3 Causalist objection

There is a causalist objection (Mele 1997) to Frankfurt’s guidance, and therefore to my proposal that guidance is necessary and sufficient for intentional action: that guidance itself depends, causally, on the agent’s psychological states; and that therefore Frankfurt fails to replace the causal view with guidance because the latter depends on the former.

[13] The reader might think that Simester’s case and mine are not analogous, because in my Jones&Black case Black does not actually intervene, while in Simester’s case Alice does. But that doesn’t matter: what matters is guidance. And in both mine and Simester’s case, both agents have guidance over the relevant movement.

Mele directs his objection to Frankfurt's already quoted car scenario. The crucial passage from that scenario, with regards to Mele's objection, is the following:

> A driver whose automobile is coasting downhill in virtue of gravitational forces alone might be satisfied with its speed and direction, and so he might never intervene to adjust its movement in any way (Frankfurt 1978, p. 48).

Mele thinks that this scenario itself depends on the attribution of psychological states:

> In the absence of a desire or intention regarding 'the movement of the automobile', there would be no basis for the driver's being 'satisfied' with the speed and direction of his car. So we might safely attribute a pertinent desire or intention to the driver, whom I shall call Al. What stands in the way of our holding that Al's acquiring a desire or intention to coast down hill is a cause of his action of coasting, and that some such cause is required for the purposiveness of the 'coasting'? … his allowing this [the 'coasting'] to continue to happen, owing to his satisfaction with the car's speed and direction, depends (conceptually) on his having some relevant desire or intention regarding the car's motion (Mele 1997, p. 9).

Here one might think that Mele's objection might apply to Frankfurt's guidance in general, but that it doesn't apply to automatic actions, and that therefore I don't need to bother with it. It wouldn't apply to automatic actions, supposedly, because I have already argued, in Chapter 3, against the attribution of psychological states in the automatic case. But I think that the causalist would have an easy reply to this move: because Mele is arguing that guidance depends on psychological states, in the absence of those psychological states, the agent's movements would not be within the agent's guidance. So what I need to argue contra Mele is, indeed, that guidance does not depend on psychological states being the causes of the agent's movement.

One clarification: here Mele is not just arguing that the agent's interventions and corrections depend, causally, upon psychological states. He is arguing that the agent's movements, even in all those cases in which no corrections or interventions take place, depend causally upon psychological states.

Mele thinks, then, that we can "safely attribute" the relevant psychological states, and that nothing stands in the way of thinking that those psychological states are causing the agent's behaviour. "Then it is natural to say that Al is coasting in his car because he wants to, or intends to, or has decided to – for an identifiable reason. And the 'because' here is naturally given a causal interpretation. In a normal case, if Al had not desired, or intended, or decided to coast, he would not have coasted; and it is no accident that, desiring, or intending, or deciding to coast, he coasts" (Mele 1997, p. 9).

But it is not enough for Mele to show that it is possible to attribute the relevant intention to the agent – namely, the agent's intention to coast. What Mele needs to show is that the attribution of the intention to coast is necessary in order for the agent to coast. If Mele doesn't show that, then he leaves room for an alternative account, one on which there is no intention to coast. It might be, for example, that all the agent intends to do is getting home: and that, because coasting doesn't undermine the satisfaction of that intention, the agent doesn't intervene. The agent's intention to get home doesn't imply the agent's intention to coast: it might be that the agent's intention to get home leaves room for the agent's intention to coast, given that coasting is, admittedly, one of many ways in which the agent can satisfy her intention to get home. But, again, that is not enough: what Mele needs is to show that the intention to coast is necessary. That, namely, the agent could not have coasted without an intention to coast; rather than just that the agent could have been coasting as the result of an intention to coast. Mele has only shown the latter, but not the former, and that is why Frankfurt's account stands.

Mele's point might show that the agent doesn't intend not to coast – because if she had intended not to coast, presumably, since her behaviour was under her guidance, she would not have coasted – but showing that the agent doesn't intend not to coast falls short of attributing any intention to the agent; and, more importantly, it doesn't show that the agent intends to coast. So that too isn't enough.

Mele is looking for a reason not to attribute psychological states to the agent; and a reason not to take them to cause the agent's movements. But what Mele needs, in order to refute Frankfurt, is to show that there cannot be guidance without those psychological states causing movement. Frankfurt's challenge is exactly that guidance doesn't depend on causal antecedents.

Because all that Mele shows is that it is possible to attribute those psychological states, Mele does not show that guidance isn't possible without those psychological states. In order to show the latter, Mele should have argued that the attribution of those psychological states is necessary, and not merely possible.

Here Mele might point out that the intervention isn't possible without the agent being in some mental state; and that if the agent is not able to intervene, then she hasn't got guidance over her actions. So guidance does depend on the agent being in some psychological state – this reply, importantly, would mean that Mele gives up on trying to show that the movements in question are caused by psychological states; and settles for just showing that the agent's capacity for guidance depends on psychological states of the agent.

But, again, all that is needed, if anything, for the agent's intervention is some intention to get home. If something happens or is about to happen that might undermine the satisfaction of such intention, then the agent might intervene. But her intervention doesn't require an intention to coast, nor does her intervention show that the agent had an intention to coast.

But I think that Mele's objection to Frankfurt doesn't work even if we grant Mele the attribution of the relevant intention – Al's intention to coast.

> Frankfurt might reply that even if Al's coasting has a suitable mental cause, his coasting his purposive 'not because it results from causes of a certain kind, but because it would be affected by certain causes if the accomplishment of its course were to be jeopardized'. The idea is that what accounts for the purposiveness of the coasting is not any feature of how it is caused but rather that Al 'was prepared to intervene if necessary, and that he was in a position to do so more or less effectively' (Mele 1997, p. 10).

Mele thinks this would be problematic for Frankfurt, and argues so, funnily enough, with a Frankfurt-type case (see section 4.2) in which the driver is under the control of a demon, such that if the driver decides to intervene, the demon will prevent him:

> Imagine that, throughout the episode, Al was satisfied with how things went and did not intervene. He decided to coast and the coasting was purposive. Imagine further that although Al intended to intervene if necessary, an irresistible mind-reading demon would not have allowed him to intervene. If Al had abandoned his intention to coast or had decided to intervene, the demon would have paralysed Al until his car run its course. The coasting is purposive even though Al was not 'in a position to [intervene] more or less effectively'. And this suggests that what accounts for the purposiveness of Al's coasting in the original case does not include his being in a position to intervene effectively (Mele 1997, p. 10).

There are two problems with Mele's argument here: he is using the conclusion he wants to defend, that Al's behaviour is purposive, as one of his premises: "He decided to coast and the coasting was purposive" (ibid). Therefore his argument is circular. Furthermore, he takes having decided to coast as the reason for the purposiveness of coasting, when that's exactly the point he has to prove contra Frankfurt's argument that the purposiveness depends, rather, on the agent's ability for guidance. Finally, it is open to Frankfurt, given what Mele says, to simply reject that the agent's movements are purposive, because the agent is not able to intervene upon them.

There is one last point that Mele makes against Frankfurt: "There are, moreover, versions of the case in which Al's coasting is purposive even though he is not prepared to intervene. Suppose Al is a reckless fellow and he decides that, no matter what happens, he will continue coasting. He has no conditional intention to intervene. Even then, other things being equal, his coasting is intentional and purposive” (Mele 1997, p. 10).

Two points: firstly, Mele is misrepresenting Frankfurt’s argument here. Frankfurt’s view, and therefore mine, relies on the agent’s ability to intervene, not on their willingness.[14] Secondly, here Mele ends up showing that purposiveness does not even depend on a conditional intention to intervene, which would have been one of the ways open to causalists to reduce guidance back to psychological states.

There could be a different objection brought against Frankfurt on a similar line as Mele’s: *that* objection might just assume some intuitive idea of purposiveness, and therefore *that* objection would not have the problems just highlighted for Mele’s objection. A proponent of that objection would then say that, in the case when the demon prevents the driver from intervening, it is intuitive that the driver’s coasting is purposive, even though the driver does not have guidance because she isn’t able to intervene.

There are various ways of dealing with such an objection: one can, as Simester (1996) proposes, slightly modify Frankfurt’s account, eliminating the part in which Frankfurt

[14] It might also rely, in fairness, on the agent’s “readiness” (Frankfurt 1978, pp. 47-48). But readiness need not be understood in terms of willingness. That an agent is ready to intervene might just mean that she is capable of doing so, that all the necessary arrangements have been made, that she can do so directly. Indeed, we can easily conceive of someone who is ready to intervene, even though she is unwilling to do so.

requires for the intervention to be 'effective': "The inability to ensure that behaviour occurs in the teeth of interfering forces does not mean that the behaviour is not action when no such forces interfere" (Simester 1996, pp. 169-170). Otherwise, as I have already mentioned, Frankfurt could just dispute the intuitiveness of the purposive character of the case: if, indeed, the agent is not in control, one might be willing to not attribute the relevant movements to the agent. Finally, one could argue, alternatively, that the sense in which the agent has, nevertheless, guidance, is that, counterfactually, had the demon not been there, she would have been in a position to intervene effectively. And that therefore, given the agent's ignorance of the demon's presence, the movements can still be attributed to her.

4.4 Explanation and Reasons

The final objection against my view that I want to deal with is that my 'guidance view' is not explanatory. This point actually contains two different objections against my view: that my view does not explain φ-ing; that my view does not provide the agent's reasons for φ-ing.

The first objection is that my 'guidance view' does not explain, causally or otherwise, why φ-ing happened. Causal theories like Davidson's, on the other hand, do. By providing the reasons – pro attitudes, beliefs (intentions in Bratman) – that motivated an agent to act, causal theories not only show that the action was intentional, but also provide a causal explanation of the action in question. And on Davidson's thesis, as we saw in Chapter 3, the causal explanation of φ-ing just is its rationalization. So Davidson gives us all three in one: the intentionality of φ-ing; the explanation of φ-ing; and its rationalization. All my view has to offer, on the other hand, is an account of the intentionality of φ-ing. It says nothing about why φ-ing happened, nor does it offer the agent's reasons for φ-ing.

But this should come as no surprise if we remember Frankfurt's original complaint against the causal view: that it focused on the antecedents of actions, rather than on the relationship between the agent and her action at the time of action. Focusing on the antecedents of action, the agent's psychological states, the causal view can offer, on top of an account of intentional action, a causal explanation and a rationalization of action. But, as we have seen, it is exactly because it focuses on the antecedents that its account comprises only of necessary conditions, renouncing to offer sufficient conditions – so that it can avoid the problem of deviant causal chains. And also, as I have argued throughout this thesis, it is because of its requirement on particular mental states causing action that it fails to account for the intentionality of automatic actions.

So offering an account of intentional action alongside a causal explanation comes at a high price for the causal view. My view, I have shown in this chapter, pays no such price: it accounts for automatic actions, and it offers necessary and *sufficient* conditions for intentionality. Also, I have only argued that the causal view fails with automatic actions; not with all kinds of actions. So it is possible that the causal view will successfully account for the intentionality of non-automatic actions. And it would then be possible to use the causal view to explain and rationalize non-automatic actions. But how do we explain and rationalize automatic actions? Let me start with explanation.

When it comes to automatic actions, I want to propose, the explanation of why an agent did something can be found in facts about the agent and/or facts about the agent's environment. For example the *fact* that the agent is very tall counts in favour of her bending when she walks through doors; and whenever she automatically bends, she bends because she is very tall. Her being very tall makes her bend. If the agent just bends, without actually intending to bend, then the explanation of why the agent bent might be that, since she has always been very tall, she was brought up to bend when walking through doors – and now she just bends every time she walks through a door.

This idea is inspired by Dancy's concept of *reasons why* (2000)[15]:

> What explains why so many people buy expensive perfume at Christmas is the barrage of advertising on the television. What explains why he didn't come to the party is that he is shy. In none of these cases are we specifying considerations in the light of which these things were done. But in all of them we are explaining why they were done. It seems, therefore, as if there is a wide range of things we think of as capable of giving answers to the question 'Why did he do that?' These answers range from specifying the things in the light of which the agent chose to do what he did, which we have sometimes called the agent's reasons for doing what he did, to something that is not a reason at all, really, but rather a cause. So we need to keep the notion of a motivating reason separate in our minds from the more general notion of 'the reason why the agent did what he did' (Dancy 2000, pp. 5-6).

[15] See also Hume's *natural instincts*. See Treatise Part III, Book III, Section IV 'Of natural abilities'. And Campbell (2006) for a discussion of how Hume's natural instincts (or natural abilities) can be reasons.

A reason why can explain why an agent did something without appealing to any mental state of the agent – facts about the agent and its environment are enough. The idea is that, when the chap in the example was asked whether he wanted to go to the party, he automatically said he wouldn't go; or anyway that, whatever his answer was, when it came to actually going to the party there was no question of going; it came *naturally* to him not to go. He didn't have to deliberate whether to go and make any decision about it; his shyness did that for him.

This kind of explanation can consist both of natural talents like being tall, or acquired talents (skills, like playing the piano), or character traits, habits, social conventions and rules. I take it that a particular talent or social convention might explain why an agent did one thing rather than another. I went to mass because I was raised as a Catholic; I didn't go to the party because I'm shy. I took that turn out of habit. One can employ both the agent's *nature* (being tall, say) and the agent's *second nature* (playing the piano wonderfully, say) to explain action. Habits, I take it, belong to this latter kind: my smoking habit explains why I couldn't resist another cigarette. My habit of listening to Radio 4 every morning explains why I listened to Radio 4 *this* morning.[16]

The fact that the guy is shy explains why he didn't go to the party, and that is independent from whether he is actually aware that he is shy or not. He needn't have any cognitive relationship with his being shy for his shyness to make him not go to the party. Clearly, agents will be aware of some facts about themselves that can count as reasons

[16] On explaining action by appealing to habits, see Pollard (2006b).

why; a tall person will normally know that they are tall (but how often will they know whether they are tall enough to bang against a door-frame? That kind of exact knowledge is rare and time-consuming, and that might very well be why tall people bend a lot of the time anyway. Bending automatically all the time might result in, sometimes, bending when it isn't necessary. But, on balance, that is probably a more effective strategy than measuring out every single doorframe). But the idea is that, anyhow, what explains their bending in going through the door is not that they *know* that they are tall, but the simple fact *that* they are tall.

So automatic actions could be explained by appealing to those facts about agents and their environments. But, it will be replied, if we appeal to those facts, we might be able to *explain* automatic actions, but we won't be able to *rationalize* them, because we are not explaining those actions *from the point of view of the agent*. If we make no mention of the considerations *in the light of which* the agent acted, then we are not in the business of rationalizing. And this is just the second objection against my view: that it does not provide the agent's reasons for φ-ing.

This point, again, I must accept. In the statement of my view there is no trace of the agent's reasons for φ-ing. But, again, it is not obvious that this should count against my view. On the other hand, here I want to show that not appealing to reasons is an advantage of my view.

The causal view, in offering the agent's reasons for φ-ing in its account of the intentionality of φ-ing, commits itself to the following, problematic, claim: that an agent acts intentionally only if she acts for a reason. This claim is at the root of the causal view's problem with automatic actions, because, as we have seen, we are not always warranted in constructing the primary reason that we need in order to claim that the automatic action in question is intentional. And this claim is also challenged by Hursthouse's (1991) *arational actions*: actions that are intuitively intentional but that cannot be rationalized by a belief-desire pair. Here are some examples given by Hursthouse: rumpling someone's hair, "throwing an 'uncooperative' tin opener on the ground" (ibid, p. 58), jumping up and down in excitement, "covering one's face in the dark [out of shame]" (ibid), "covering one's eyes [in horror] when they are already shut" (ibid).

So there are at least two kinds of actions, automatic actions and arational actions, that my view, differently from the causal view, can account for: and that is exactly because my view does not appeal to reasons. But while we could easily conclude that arational actions are not done for a reason, we don't want to say that all automatic actions are like that too. When I automatically flip the switch, or when I bend in walking through a door - differently from when I jump up and down in excitement - my behavior is goal-directed and rational; and I normally have a reason for doing it.

All I have been questioning in this thesis is, after all, that I need have, in every case, a psychological state in the shape of an unconscious belief that flipping the switch will satisfy my pro attitude; because without such belief my flipping the switch would not be intentional. This does not mean, evidently, that I could not have had a reason for flipping the switch.

But what kind of account can I offer of the agent's reasons in automatic cases? I can't appeal to psychological states. Take again the turning off the light example. If I said that I turned off the light because I want to reduce my carbon footprint, and took my *desire* to 'reduce my carbon footprint' as the reason for my turning off the light, that would seem a perfectly sensible rationalization of my behavior. The problem is that the Davidsonian would reply that 'reducing my carbon footprint' can rationalize my 'turning off the light' *only* if I believed that 'turning off the light' had the property of 'reducing my carbon footprint'. So I need, after all, the belief against which I have been arguing in Chapter 3.

One could then try and say that the facts about the agent and her environment might be her reasons. So that the fact that I am very tall will be the reason for my bending when I walked through my office door. But, it will be objected, this is no rationalization because I am not including the point of view of the agent; the fact that I am very tall makes no mention of the considerations in the light of which I acted.

But, it might be replied, facts can't be reasons only if one accepts Davidson's internalism, according to which reasons must be psychological states of the agent. According to externalists (such as Stout 1996, Collins 1997, Dancy 2000), on the other hand, facts can rationalize an agent's behavior. So that if I bend because I am very tall, what rationalizes my bending isn't my *belief* that I am very tall, but the *fact* ("objective circumstance" Collins 1997, p. 109) that I am very tall.

Externalists admit, on the other hand, that the fact that I am very tall can rationalize my action of bending only if I have some grasp of the fact that I am very tall (see Dancy 2000, Ch. 1; also Stout: "So my denial of the Internalist Shift does not involve me in denying that an agent must have mental access to the immediate reasons for their actions" (1996, p. 38)). So in order to construct an externalist rationalization for automatic actions, one mustn't end up cashing out this "mental access" requirement so as to violate my lack of attention and awareness condition on automatic action.

The alternative is giving up on the assumption that my pro attitude towards 'reducing my carbon footprint' can rationalize my turning off the light *only* if I also have the belief that *'turning off the light' has the property of 'reducing my carbon footprint'*. Gert (1998) offers one such solution: for an action to be rational, it suffices that it is not irrational. "Defining a rational action simply as an action that is not irrational does not impose a fictitious and misleading uniformity on all rational actions" (1998, p. 61). So if our agent doesn't have any reasons against bending, then her bending is rational just in virtue of the fact that the agent has no reason not to bend. Here, what rationalizes the action is the absence of reasons rather than their presence. But since we, differently from the causalist, are not committed to reasons being causes, this is not necessarily a problem for us.

Gert's proposal appears to fit automatic actions particularly well. In Chapters 3 and 4 I have not argued that agents, when acting automatically, have no intentions, beliefs, or desires at all. I have only argued against the attribution of particular intentions or beliefs. Suppose I am walking to work: I will take most of my steps automatically. Now what I have been arguing is that, in order for an individual step to be an intentional action of mine, I don't need to have a belief that *that* particular step has the property of taking me to work.

Here Gert offers us the opportunity to say a similar thing about rationality: I don't need a belief with the relevant content in order to make my taking *that* step rational, as long as there is nothing that makes it irrational for me to take that step (as in, for example, nothing that is inconsistent with my taking that step). So that 'going to work' can be the reason for my taking that particular step independently from my having the relevant belief, as long as I don't believe that taking that particular step will interfere with my 'going to work' That is, as long as some automatic action does not interfere with my overall plans, I let myself do it. And it is perfectly rational to do so.[17] At least for rationality, then, there might be a solution to the gap between ψ-ing (walking to work) and φ-ing (taking *that* individual step) that Bratman failed to fill with motivational potential: leaving it blank.

[17] This proposal, then, would not only offer a way to rationalize automatic actions, but also a way to *justify* them.

Conclusion

In this chapter I have presented an account of intentional action according to which automatic actions are intentional, the 'guidance view'. This account is a development on Frankfurt’s idea that guidance is sufficient in order to distinguish between actions and mere bodily movements. I argue that guidance is also sufficient for intentional action. I show that this elaboration on Frankfurt still enables me to distinguish between when agents act intentionally and when they don’t – and so I can still allow for what the causal view calls unintentional actions. Also, I have shown that my proposal has a major advantage over the causal view: I can give necessary and sufficient conditions for intentional action, while the causal view can only give necessary conditions. This is because my view can account for deviant cases as cases of mere bodily movements. In the rest of the chapter, I have dealt with four potential objections: that my view is subject to regress; that my view is subject to a version of Frankfurt’s own counterexample against the Principle of Alternative Possibilities; that guidance can be reduced back to causal views; and that my proposal does not explain nor rationalize action.

Conclusion

I have already stated my conclusions: I have said why I don't think that the causal views which I have analysed – Davidson's, Bratman's, and the Simple View – work in the case of automatic actions. And I have proposed an alternative account, my 'guidance view'. Therefore I consider my argument complete, and I don't see the need to summarize it here.

So what I want to do in this Conclusion is only to highlight a particularly interesting consequence of my account: the simplification of the relationship between intentionality and responsibility, so that agents are responsible for all and only their intentional actions.

Someone who accepted the causal view would have to concede that at least some of the actions that the causal view would consider unintentional actions are actions for which agents are responsible. Take Davidson's (1978) famous Bismarck scenario, where the officer in charge of the torpedoes mistakenly sinks the Bismarck thinking that it is the Tirpitz.[1] The idea is that the officer sinks the Bismarck unintentionally; because his[2] actions, under the description 'sinking the Bismarck', are not rationalized: he does not have a pro attitude towards 'sinking the Bismarck', nor does he have a relevant belief that, together with his pro attitude towards 'sinking the

[1] Interestingly enough, Davidson's scenario is historically inaccurate. When the Royal Navy sank the Bismarck, the Tirpitz was nowhere near.

[2] I don't think there were any female officers aboard Royal Navy battleships during World War II; that's my reason for choosing to use the male pronoun.

Tirpitz', would rationalize 'sinking the Bismarck': something like the belief that his action, under description 'sinking the Bismarck', would have the property of 'sinking the Tirpitz'. So 'sinking the Bismarck' is, under *that* description, an unintentional action of the agent. The officer has, in short, made a mistake.

Whether the officer ought to be held responsible for his mistake is a different matter, and one that Davidson does not discuss. But it is a matter of interest to us: so let us suppose that the officer is, indeed, responsible. We could suppose that there is a standard procedure to identify enemy battleships, and that had the officer followed such procedure, he would have easily identified the battleship as the Bismarck rather than the Tirpitz, thereby avoiding the mistake. So the officer is responsible for his mistake because he did not follow standard Navy procedures; and because, we are supposing, following such procedures would have meant avoiding the mistake.[3]

So the officer is responsible for 'sinking the Bismarck' even though he did so unintentionally. Therefore some unintentional actions are still actions for which agents are responsible. This is a quite familiar idea: cases of *ignorance* and *negligence* (see, for example, Rosen 2001 and 2003) are, for example, cases of unintentional actions for which the agent is nevertheless responsible. Those are cases in which the classic reply "I didn't do it intentionally" (or: "I didn't mean to do it") is no excuse from responsibility. The Bismarck scenario would then be a case of negligence, because the agent failed to follow standard procedures.

[3] One might think that the truth of the counterfactual "had the officer followed procedures, the mistake would have been avoided" is not necessary for the officer's responsibility, because the simple fact that he did not follow procedures is sufficient. This point is unsubstantial just now: what matters is that we are supposing that the officer is responsible, for whichever reason.

On my view, on the other hand, there are no cases of unintentional actions for which the agent is nevertheless responsible: if the agent acts intentionally, then she is responsible for what she has done. If she does not act intentionally, then she is not responsible for what she has done. This is because on my view intentionality depends on whether the agent had guidance over her movements, and on whether the agent can be expected to know or find out some description of her movements at no unreasonable costs (which I have called ‘guidance over her actions’ as opposed to ‘guidance over her movements’). If the agent doesn’t have guidance over her movements, or if the agent does have guidance but can’t be expected to know or find out some description of her movements at no unreasonable cost, then she does not act, under that description, intentionally. This means that I, differently from the causal view, don’t need to allow for the possibility of the agent being responsible for something that she doesn’t do intentionally.

Let us look at the Bismarck scenario again. On my view, the agent sinks the Bismarck intentionally. Not only did the agent have guidance over her movements, because she could have at any time directly intervened to either stop or redirect the launch of the torpedo. But also the agent can be expected to know or find out at no unreasonable cost that the battleship at which he is firing is not the Tirpitz. Indeed, what it will take the officer to find out is *only* following standard procedures – which is his duty anyway. Therefore it is not unreasonable to expect that he follow such procedures, thereby finding out that the battleship is the Bismarck rather than the Tirpitz.

What this difference between the ‘guidance view’ and causal views means is, in short, that my view offers grounds for holding agents responsible for their actions, while the causal view doesn’t. On my view, the agent is responsible because she acted intentionally – acting intentionally is, on my view, both necessary and sufficient to being responsible for your actions. On the causal view, on the other hand, the agent does not need to act intentionally to be responsible, as in the Bismarck scenario. But then what is it, on the causal view, that makes an agent responsible?

It is not that she acted unintentionally either, because obviously we would not want to say that an agent was responsible for all her unintentional actions. Suppose you are passing me a cup of coffee, and an earthquake causes you to spill the coffee on my skirt. Now, you have spilled the coffee unintentionally, but it would be unreasonable to hold you responsible for it, given that you could not have resisted, we are supposing, the earthquake. So not all unintentional actions are actions for which the agent is responsible; and not all intentional actions are actions for which the agent is responsible either. Acting intentionally is therefore not necessary for responsibility (it might be that it is sufficient, but I won’t go into that). So the causal view does not offer grounds for responsibility.

There is a possible reply that the causalist could offer to the earthquake scenario. The causalist could argue that, at the very moment in which the earthquake causes you to spill the coffee, there is no description under which your action is intentional, and that therefore spilling coffee is not an unintentional action, because it isn't an action at all – it's just a way in which your body has been caused to move by the earthquake. So it's not that you spilled coffee, but that the coffee has been spilled because of the earthquake (you are just a proximal cause of the 'coffee spilling' event, but not its agent).

The disagreement here is about whether, once the earthquake's causal influence on you has initiated, it is still the case that your movements are intentional under the description 'offering coffee'. And to my proposal that they are, the causalist replies that they aren't: that you are no longer offering me coffee once your movements are being influenced by the earthquake. This appears counterintuitive: if an observer were to describe the scene, she would probably say that, while you were passing me the coffee, you spilled some. Also, because the causalist's criterion is whether some movement was caused by a particular mental state, 'spilling coffee' meets such criterion even in the earthquake scenario, because the relevant mental state is, despite the earthquake, one of its causes.

Here it looks as though a causalist who is committed to Davidson's understanding of action individuation will have to concede that 'offering coffee' and 'spilling coffee' are two different descriptions of the same action. If, indeed, the causalist were willing to give up on such commitment, then they might have a way of saying that all unintentional actions are actions for which agents are responsible, because those for which they are not responsible are not even actions – which is exactly my view.

But then the causalists would have come a long way towards the side of guidance, in that they would have recognized, crucially, that what distinguishes actions from mere movements is not their preceding mental states, but rather whether the agent has appropriate control over the way in which her body moves: the appropriate control that Frankfurt and I call guidance.

Let me clarify my claim here: that the causal view does not offer grounds for responsibility is no objection against the causal view, because it could be replied, quite fairly, that responsibility is outside the scope of such a view. My claim here is only that my view, which should be preferred to the causal view for reasons already stated in this thesis, has a further advantage over the causal view: it offers necessary and sufficient grounds for holding an agent responsible for some action: *E is responsible for φ-ing iff E φ-ed intentionally*.

Obviously this is no more than a sketch of what a full account of responsibility should look like. To complete it, it will take specifying what it is reasonable to expect of an agent in each particular situation; and what, in each particular situation, is an unreasonable cost. Also my view, it must be emphasized, does not offer necessary and sufficient conditions for when an agent is responsible for something (some event, say); but only necessary and sufficient conditions for when an agent is responsible for some action. And I have said nothing about whether her own actions are the only events for which an agent is ever responsible.

Here a difficult case is represented by the drunken driver scenario (see, for example, Fisher and Ravizza 1998). Suppose I go out drinking and get very drunk. Suppose that, notwithstanding my being very drunk, I drive back home. And suppose that I run over a pedestrian on my way back home. It appears quite obvious that I must be, at some level, responsible for running over the pedestrian. On the other hand, it will be difficult to argue that I intentionally run over the pedestrian. It will be difficult not only for the causal view, since I quite obviously, we can suppose, didn't intend nor had any reason to run over the pedestrian. But it will also be difficult on the 'guidance view' to claim that 'running over the pedestrian' was an intentional action of mine; because it looks as though my running over the pedestrian was a consequence of my lack of control over the car: the fact, in short, is that my being very drunk made it more difficult, if not impossible, for me, to *guide* the car.

So it seems that at least some of those cases of drunkenness are cases in which the drunken agent does not have guidance over her movements; she cannot, therefore, be acting intentionally. Here there are therefore two possibilities in arguing for the drunken agent's responsibility over what she does when drunk. On the one hand, it could be argued that what she does cannot count as her actions, because of lack of control or guidance, and that we must therefore admit that there are events other than the agent's own intentional actions for which an agent will be held responsible – so that the case of 'running over the pedestrian' will be one such event.

But I think that we don't necessarily need to go down that route: guidance might still provide us with a way to argue for the agent's responsibility for 'running over the pedestrian’. The idea is that the agent did have guidance over whether to go to the pub; she did have guidance over whether to *drive* to the pub (rather than taking the bus, say); she did have guidance over whether to drive back home; and so on. In short, we can find a lot of previous actions of the agent which are intentional. Those actions, as it happens, lead to the agent’s running over the pedestrian.

Furthermore, it looks as though it would have been reasonable to expect that the agent had known, at the time when she decided to drive to the pub, that getting very drunk might have resulted in having to drive home drunk, and that driving home drunk might have resulted in an accident – and that the agent would have known or could have found out those possible consequences of her actions at no unreasonable cost. Those are the kinds of consequences that a person *should* expect from drinking and driving. So it might be that those are sufficient grounds for holding the agent responsible for doing something over which, at the time of action, she had no control (for an similar account of *historical* control and responsibility see Wright 1976).

As I already stated, I do not pretend to have given a complete account of responsibility. But I hope to have shown that the concept of guidance, and its application to intentional action, are very promising in developing a full account of responsibility.

Appendix

The results of the two surveys reported below show that the intuition that automatic actions are intentional is widely shared.

1. First survey

I interviewed 50 subjects in Edinburgh in June 2007: they were mostly university students. The experiment consisted in them reading a short story, and then answering four questions about the story they had just read. They were not allowed to look at the questions before reading the story; nor were they allowed to look at the next question before having answered the previous one. The questions were answered always in the same order. Subjects were allowed, though, to look back at the story when answering a question.

The story went as follows:

Sarah was sitting on her bed, desperately hoping for Mark to call. Staring at her phone like in the movies, Sarah was thinking how wonderful it would be to hear his voice again. She got up and went over to the window, relishing the prospect of one of those long conversations with Mark. Then the phone rang, Sarah answered: it was him!

The four questions were the following, and had to be answered always in the same order as they are presented below:

1) *Did Sarah intend to get up?*

2) *Did Sarah get up intentionally?*

3) *Did Sarah intend to answer the phone?*

4) *Did Sarah answer the phone intentionally?*

The idea is that Sarah ‘gets up’ automatically, without thinking about it (possibly without even noticing that she does so). On the other hand, ‘answering the phone’ is something that Sarah has long anticipated, something that she has given a lot of thought to: an action, in short, that we could hardly imagine to be automatic.

The answers were as follows:

Question	*NO*	*YES*	*Don’t know/no answer*
1	20	29	1
2	11	39	/
3	5	43	2
4	6	41	3

These answers amount to the following percentages:

Question	*NO*	*YES*
1	40%	58%
2	22%	78%
3	10%	86%
4	12%	82%

1.1 Discussion

I have picked a spontaneous automatic action, such as ‘getting up’ absentmindedly, and a deliberated non-automatic action, such as ‘answering the phone’ for a much awaited phone call.

I wanted to test two things: whether subjects were as willing to attribute an intention to the agent for the automatic action as for the non-automatic action. And, more importantly, whether subjects’ intuitions about the intentionality of the two actions were different.

It emerges that subjects, in their attributions of intentions, acknowledge the difference between the automatic action of ‘getting up’ and the non-automatic action of ‘answering the phone’: while only 10% are unwilling to attribute an intention in the latter case, as much as 40% answered that the subject did not intend to get up.

On the other hand, subjects don’t appear to distinguish between the two actions in terms of intentionality: 78% considered ‘getting up’ intentional, and a very similar 82% considered ‘answering the phone’ intentional.

I think that those results lend the support of people’s intuitions to my claim that there is a relevant difference between automatic actions and non-automatic actions in terms of their preceding mental states: I have argued that causal views, because they rely on the attribution of mental states in every case, fail to account for automatic actions. Also these results, most importantly, confirm the intuition behind my thesis, that automatic actions are intentional.

1.2 Philosophers vs laypeople

Since the survey was about philosophical intuitions, I have also recorded whether the subject was a philosopher or a layperson; where 'philosopher' was defined as someone who was at least doing a postgraduate (Masters/PhD) course in philosophy. 17 respondents were philosophers, 33 were laypeople.

Here is the breakdown:

Laypeople:

Question	*NO*	*YES*	*Don't know/no answer*
1	12	21	/
2	7	26	/
3	4	29	/
4	3	28	2

Philosophers:

Question	*NO*	*YES*	*Don't know/no answer*
1	8	8	1
2	4	13	/
3	1	14	2
4	3	13	1

These answers amount to the following percentages:

Laypeople:

Question	*NO*	*YES*
1	36.36%	63.64%
2	21.21%	78.79%
3	12.12%	87.88%
4	9.09%	84.85%

Philosophers:

Question	*NO*	*YES*
1	47.06%	47.06%
2	23.53%	76.47%
3	5.88%	82.35%
4	17.65%	76.47%

The most striking difference between the general results and the specific results of the two categories is certainly the philosophers' answer to question 1, which was evenly split: 47% did not attribute an intention to 'get up', and another 47% did. This shows that philosophers are apparently less willing than average to attribute an intention in the case of automatic actions: 47% against the general 58% and the laypeople's 64%.

2. Second survey

The second survey pursued the same two hypotheses through different stories, and with a different methodology.

This time, the automatic action and the non-automatic action did not feature in the same story, but in two distinct sketches. Also, this time the automatic action and the non-automatic action were of the same kind, ‘boiling the kettle’; while in the first survey they were of different kinds, ‘getting up’ and ‘answering the phone’.

Below are the two stories. In #1, ‘boiling the kettle’ is supposed to be an automatic action, as part of the agent’s morning routine; something that she habitually does every morning.

#1:

Today Karen woke up with the unpleasant consciousness that she had an interview. She had taken ages to fall asleep the night before, and still now she could think of nothing else: should I wear a skirt or trousers? Should I walk or get a taxi? All the same, she got on with her usual morning routine: she opened the shutters, then went to the kitchen, boiled the kettle, turned on the radio, and sat down for her breakfast.

On the other hand, in #2 ‘boiling the kettle’ is something that the agent does after having resolved a dilemma over whether to do it or not; therefore she doesn’t do it automatically. She has actually had to think about it.

#2

Karen couldn't decide whether to have espresso or instant coffee. Espresso, she thought, tastes nicer. But with instant you get more, much more. As always with her, quantity prevailed over quality, and she decided to put on the kettle for a big cup of instant coffee. So she boiled the kettle.

For both stories the same pair of questions were asked:

1) *Did Karen intend to boil the kettle?*
2) *Did Karen boil the kettle intentionally?*

The methodology, for this second survey, was very different. No subject was shown both stories, and no subject was asked more than one question about either story. So there were four sets of subjects, each answering only one question. This was done in case, in the first survey, answers to later questions had been influenced by answers already given by the same subject to earlier questions. On the other hand, while the methodology of the first survey tested a person's contrastive intuitions about intention and intentionality, this second methodology does not, because each subject answers only one question.

For this second survey, conducted over the internet, I have interviewed 357 people: one hundred each for each question of story #2; one hundred for question 1 of story #1; and 57 for question 2 of story #1.

The results were as follows:

Story #1:

Question	*NO*	*YES*
1	29	71
2	11	46

Story #2:

Question	*NO*	*YES*
1	42	58
2	6	94

These results amount to the following percentages

Story#1:

Question	*NO*	*YES*
1	29%	71%
2	19.3%	80.7%

Story #2:

Question	*NO*	*YES*
1	42%	58%
2	6%	94%

2.1 Discussion

The results of the second survey confirm, importantly, that the intuition that automatic actions are intentional is widely shared: an overwhelming majority of 81% of the respondents said that, in the automatic case, Karen intentionally boiled the kettle. Similarly, an overwhelming majority of 94% said that Karen boiled the kettle intentionally in the non-automatic case in which she does it as the result of a dilemma.

On the other end, the results of this second survey do not confirm the first survey as to people's unwillingness to attribute intentions in the automatic case as opposed to the non-automatic case: 58% of respondents attribute an intention in the non-automatic case, and even more, 71%, in the automatic case.

There is an obvious way to explain the diverse findings of the two surveys on the attribution of intention: while in the first survey 'getting up' was presented as a spontaneous automatic action, in the second survey 'boiling the kettle' is presented as an habitual automatic action which is part of a routine: it is therefore obviously goal-directed. And, crucially, the goal is presented to the respondent in the story in a way in which the first survey didn't do: there, the explanation of why Sarah had gotten up was somehow left to the respondent. While here a respondent can be in no doubt over what are Karen's goals.

There is another important consideration to make about those results on the attribution of intention. There is a crucial difference, as long as the philosophy of action is concerned, between not intending to 'boil the kettle' and intending to 'not boil the kettle' – where, crucially, the negation is part of the content of the intention only in the latter case. The difference is that in the former case no intention is attributed to the agent. And so the former case could not be cited as part of a causal explanation as those of Bratman and Davidson, because the mental state that is supposed to have caused action is missing: to say that the agent does not intend to boil the kettle does not attribute any intention to the agent.

But it is not obvious that this difference is picked up on by non-philosophers: after all, it isn't obvious that non-philosophers are committed to Davidson's and Bratman's causalism. Indeed, it might be that, in these surveys, the attribution of intention is explained by the fact that respondents are mostly concerned with not saying that the agent's intentions are against 'getting up' or 'boiling the kettle'. That much appears obvious from the stories: that the agents are not against 'getting up' and 'boiling the kettle'. But that does not mean, yet, that they actually intend to do those things.

It might be that respondents shy away from not attributing the intention to make sure that they do not end up saying (or anyway being taken to say) that the agent's intentions and attitudes were against 'getting up' and 'boiling the kettle', which would be an obvious mistake.

But here I am only speculating: the only thing we can conclude, from the data, is that the second survey confirms only one of the two hypotheses supported by the first survey: that intuition tells us, overwhelmingly, that automatic actions are intentional. The second hypothesis - that people distinguish between automatic actions and non-automatic actions in terms of the attribution of mental states - is not supported by this second survey.

Bibliography

Adams, F. (1986), ‘Intention and Intentional Action: The Simple View’, *Mind & Language* 1: 281-301.

Adams, F. and Mele, A. (1989), 'The Role of Intention in Intentional Action', *Canadian Journal of Philosophy* 19: 511-31.

Aglioti, S. & al. (1995), ‘Size contrast illusions deceived the eye but not the hand’, *Current Biology* 5: 679-85.

Alston, W. (1986), 'An Action-Plan interpretation of purposive explanations of actions', *Theory and Decision* 20: 275-299.

Anscombe, G.E.M. (1957), *Intention*. Basil Blackwell.

Aristotle, *Nicomachean Ethics* (1925 Ross’s translation). Oxford UP.

Audi, R. (1973), ‘Intending’, *Journal of Philosophy* 70: 387-402.

Bargh, J.A. & Chartrand, T.L. (1999), ‘The Unbearable Automaticity of Being’, *American Psychologist* 54: 462-479.

Bargh, J.A., Chen, M., & Burrows, L. (1996), ‘Automaticity of Social Behavior: Direct effects of trait construct and stereotype activation on action’, *Journal of Personality and Social Psychology* 71: 230-244.

Beardsley, M. (1978), ‘Intending’, in Goldman, A. & Kim, J. (eds.), *Values and Morals*. Dordrecht.

Beilock, S. L., Wierenga, S.A., & Carr, T.H. (2002). ‘Expertise, Attention, and Memory in Sensorimotor Skill Execution’, *Quarterly Journal of Experimental Psychology* 55: 1211-1240.

Bermudez, J. (1995), ‘Nonconceptual Content: From Perceptual Experience to Subpersonal Computational States’, *Mind and Language* 10: 333-369.

Bishop, J. (1989), *Natural Agency. An Essay on The Causal Theory of Action*. Cambridge University Press.

Block, N.J. and Fodor, J.A. (1972), ‘What Psychological States Are Not’, *Philosophical Review* 81: 159-181.

Brand, M. (1984), *Intending and Acting*. MIT Press.

Bratman, M. (1984), 'Two Faces of Intention', *Philosophical Review* 93: 375-405.

Bratman, M. (1987), *Intention, Plans, and Practical Reason*. Cambridge, Mass.: Harvard University Press.

Campbell, T. (2006), 'Human Philosophy: Hume on Natural Instincts and Belief Formation', in Di Nucci, E. & McHugh, C. (eds.), *Content, Consciousness, and Perception*. Cambridge Scholars Press.

Carruthers, P. (1996), *Language, Thought and Consciousness: An Essay in Philosophical Psychology*. Cambridge: Cambridge University Press.

Chisholm, R. (1966), ‘Freedom and Action’, in Lehrer, K. (ed.) *Freedom and Determinism*. Random House.

Chisholm, R. (1966), *Theory of Knowledge*. Englewood Cliffs.

Clark, A. (2001), ‘Visual Experience and Motor Action: Are the Bonds too tight?’, *Philosophical Review* 110: 495-519.

Collins, A. W. (1997), ‘The psychological reality of reasons’, *Ratio*, X: 108-123.

Cooper, R. & Shallice, T. (2000), ‘Contention Scheduling and the Control of Routine Activities’, *Cognitive Neuropsychology* 17(4): 297-338.

D'Arcy, E. (1963), *Human Acts: An Essay in their Moral Evaluation*. OUP.

Dancy, J. (2000), *Practical Reality*. Oxford UP.

Darley, J.M. and Schultz, T.R. (1990), 'Moral Rules: Their Content and Acquisition', *Annual Review of Psychology* 41, 525–56.

Davidson, D. (1963), 'Actions, Reasons, and Causes', *Journal of Philosophy* 60: 685-700.

Davidson, D. (1969), 'The Individuation of Events', in Rescher, N. & Reidel, D. (eds.), *Essays in Honour of Carl G. Hempel*, Reidel Publishing Company.

Davidson, D. (1971), 'Agency', in Binkley, R., Bronaugh, R., and Marras, A. (eds.), *Agent, Action, and Reason*. University of Toronto Press.

Davidson, D. (1973), 'Freedom to Act', in Honderich, T. (ed.), *Essays on Freedom and Action*. Routledge and Kegan Paul, 137-56.

Davidson, D. (1978), 'Intending', in Yovel, Y. (ed.), *Philosophy of History and Action*. The Magnes Press, The Hebrew University.

Davidson, D. (1980), *Essays on Actions and Events* (2nd ed: 2000). Oxford UP.

Davis, L. (1970), 'Individuation of Actions', *Journal of Philosophy* 67: 524-5.

Davis, L. (1979), *Theory of Action*. Prentice-Hall.

Davis, W. (1984), 'A causal theory of intending', *American Philosophical Quarterly* 21: 43-54.

Davis, W. A. (1982), 'A Causal Theory of Enjoyment', *Mind* 91: 240-256.

Dennett, D.C. (1969), *Content and Consciousness*. Routledge & Kegan Paul.

Dennett, D.C. (1991), *Consciousness Explained.* Penguin.

Dretske, F. I. (1969), *Seeing and Knowing*. Chicago.

Dreyfus, H. (1988), 'The Socratic and Platonic Bases of Cognitivism', *AI & Society* 2: 99-112.

Dreyfus, H. (2005), 'Overcoming the Myth of the Mental: How Philosophers Can Profit from the Phenomenology of Everyday Expertise', *APA Pacific Division Presidential Address*.

Dreyfus, H. & Dreyfus, S. (1984), 'Skilled Behavior: The Limits of Intentional Analysis', in Lester, E. (ed.), *Phenomenological Essays in Memory of Aron Gurwitsch*. The University Press of America.

Fischer, J.M. (1994), *The Metaphysics of Free Will.* Blackwell.

Fisher, J.M. and Ravizza, M. (1998), *Responsibility and Control*. Cambridge UP.

Foot, P. (1967), 'The Problem of Abortion and the Doctrine of Double Effect', *Oxford Review* 5: 5-15.

Foot, P. (1985), 'Morality, Action, and Outcome', in Honderich, T. (ed.), *Morality and Objectivity.* Routledge & Kegan Paul.

Frankfurt, H. (1969), 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy* 66: 829-839.

Frankfurt, H. (1978), 'The Problem of Action', *American Philosophical Quarterly* 15: 157-162.

Garcia, J.L.A. (1990), 'The Intentional and the Intended', *Erkenntnis* 33: 191-209.

Gert, B. (1998), *Morality: its nature and justification*. Oxford UP.

Gert, J. (2003), 'Brute Rationality', *Nous* 37: 417-446.

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

Ginet, C. (1996), 'In Defence of the Principle of Alternate Possibilities: Why I Don't Find Frankfurt's Arguments Convincing', *Philosophical Perspectives* 10: 403-417.

Goldie, P. (2000), 'Explaining expressions of emotions', *Mind* 109: 25-38.

Goldman, A. (1970), *A Theory of Human Action*, Englewood Cliffs.

Goldman, A. (1971), 'The Individuation of Action', *Journal of Philosophy* 68: 769-

72.

Grice, H. P. (1971), ‘Intention and Uncertainty’, *Proceedings of the British Academy* 57: 263-79.

Hacker, P.M.S. & Bennett, M.R. (2003), *Philosophical Foundations of Neuroscience*. Blackwell.

Hampshire, S. (1959), *Thought and Action*. Chatto and Windus.

Hare, R.M. (1952), *The Language of Morals*. Oxford UP.

Harman, G. (1965), ‘The Inference to the Best Explanation’, *Philosophical Review* 74: 88-95.

Harman, G. (1976), ‘Practical Reasoning’, *Review of Metaphysics* 29: 431-463.

Heil, J. and Robb, D. (2003), ‘Mental Properties’, *American Philosophical Quarterly* 40: 175-196.

Hornsby, J. (1980), *Actions*. Routledge & Kegan Paul.

Hurley, S. (1989), *Natural reasons*. OUP.

Hursthouse, R. (1991), 'Arational Actions', *Journal of Philosophy* 88 (2): 57-68.

James, W. (1890), *The Principles of Psychology*. London: Macmillan.

Jeannerod, M. (2003), ‘Consciousness of Action and Self-Consciousness: A Cognitive Neuroscience Approach’, in Roessler, J. & Eilan, N. (eds.), *Agency and Self-Awareness*. Oxford: Clarendon Press.

Jeannerod, M. & Fourneret, P. (1998), ‘Limited conscious monitoring of motor performance in normal subjects’, *Neuropsychologia* 36: 1133-40.

Kelly, S. and Knobe, J. (unpublished), 'Can one act for a reason without acting intentionally?'.

Kenny, A. (1989), *The Metaphysics of Mind*. Clarendon Press.

Knobe, J. (2003), ‘Intentional Action and Side Effects in Ordinary Language’, *Analysis* 63: 190-193.

Knobe, J. (2005), ‘Theory of Mind and Moral Cognition: Exploring the Connections’, *Trends in Cognitive Science* 9: 357-359.

Knobe, J. (forthcoming), ‘The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology’, *Philosophical Studies*.

Lewis, D. (1986), ‘Causal Explanation’, *Philosophical Papers*, ii. Oxford UP.

Lewis, D. (1990), ‘What Experience Teaches’ in Lycan, W. (ed.), *Mind and Cognition: A Reader*. Blackwells.

Lhermitte, F. (1983), ‘'Utilisation behaviour' and its relation to lesions of the frontal lobes', *Brain* 106: 237–255.

Libet, B. (1985), ‘Unconscious cerebral initiative and the role of conscious will in voluntary action’, *Behavioural and Brain Science* 8: 529-566.

Lowe, J. (1999), ‘Self, Agency, and Mental Causation’, *Journal of Consciousness Studies* 6: 225-39.

Lycan, W. G. (1973), ‘Inverted Spectrum’, *Ratio* 15: 315-9.

MacIntyre, A. (1957), *The Unconscious*. Routledge & Kegan Paul.

Macrae, C.N. & Johnston, L. (1998), ‘Help, I Need Somebody: Automatic Action and Inaction’, *Social Cognition* 16: 400-417.

Malcolm, N. (1968), ‘The Conceivability of Mechanism’, *Philosophical Review* 77: 45-72.

Marcel, A. J. (1998), ‘Blindsight and shape perception: deficit of visual consciousness or of visual function?’, *Brain* 121: 1565-88.

"Marcel, A. J. (2003), 'The Sense of Agency: Awareness and Ownership of Action', in Roessler, J. and Eilan, N. (eds.), *Agency and Self-Awareness*. Oxford UP.

Maudsley (1873), *Physiology of Mind* (quoted in James 1890).

Mayr, E. (1976), *Evolution and the Diversity of Life*. Harvard UP.

McCann, H. (1986), 'Rationality and the Range of Intention', *Midwest Studies in Philosophy* 10: 191-211.

McCann, H. (1991) 'Settled Objectives and Rational Constraints', *American Philosophical Quarterly* 28: 25-36.

McCann, H. (1998), *The Works of Agency*. Cornell UP.

McDowell, J. (1978), 'Are Moral Requirements Hypothetical Imperatives?', *Proceedings of the Aristotelian Society, Supplementary Volume* 52: 13-29.

McDowell, J. (1979), 'Virtue and Reason', *The Monist* 62: 331-50.

McDowell, J. (1992), 'Meaning and Intentionality in Wittgenstein's Later Philosophy', in French, P.A., Uehling, T.E. Jr., and Wettstein, H.K. (eds.), *The Wittgenstein Legacy. Midwest Studies in Philosophy 17*. University of Notre Dame Press, pp. 40-52.

McDowell, J. (1994), 'The Content of Perceptual Experience', *Philosophical Quarterly* 44(5) [175]:190-205.

McDowell, J. (1994), *Mind and World* (with a new introduction by the author: 1996). Harvard UP.

McDowell, J. (1998), *Mind, Value, and Reality*. Harvard UP.

Mele, A. (1988), 'Intentions, Plans, and Practical Reason', *Mind* 97: 632-634.

Mele, A. (1992), *Springs of Action*. Oxford UP.

Mele, A. (1997), *Philosophy of Action*. Oxford UP.

Mele, A. and Moser, P. K. (1994), 'Intentional Action', *Nous* 28: 39-68.

Milner, D. & Goodale, M. (1995), *The visual brain in action*. Oxford: Oxford University Press.

Moran, R. (2001), *Authority and Estrangement: an Essay on Self-knowledge*. Princeton University Press.

Morton, A. (1975), 'Because He Thought He Had Insulted Him', *Journal of Philosophy* 72: 5-15.

Nadelhoffer, T. (2006), 'On Trying to Save the Simple View', *Mind & Language* 21: 565-586.

Nagel, T. (1970), *The Possibility of Altruism*. Oxford UP.

Nagel, T. (1986), *The View from Nowhere*. Oxford UP.

Norman, D.A. & Shallice, T. (1986), 'Attention to Action: willed and automatic control of behaviour', in Davidson, R.J., Schwartz, G.E. & Shapiro, D. (eds.), *Consciousness and Self-Regulation*, iv. New York: Plenum, 1-18.

Norman, R. (2001), 'Practical Reasons and the Redundancy of Motives', *Ethical Theory and Moral Practice* 4: 3-22.

Nozick, R. (1993), *The Nature of Rationality*. Princeton UP.

O'Brien, L. (2003), 'on Knowing One's Own Actions', in Roessler, J. and Eilan, N. (eds.), *Agency and Self-Awareness*. Oxford UP.

O'Shaughnessy, B. (2003), 'The Epistemology of Physical Action', in Roessler, J. and Eilan, N. (eds.), *Agency and Self-Awareness*. Oxford UP.

Pashler, H.E. (1998), *The Psychology of Attention*. MIT Press.

Peacocke, C. (1995), 'Conscious Attitudes, Attention, and Self-Knowledge', in Wright, C., Smith, B.C., and Mcdonald, C. (eds.), *Knowing Our Own Minds*. Oxford UP.

Peacocke, C. (2003), ‘Action: Awareness, Knowledge, and Ownership’, in Roessler, J. and Eilan, N. (eds.), *Agency and Self-Awareness*. Oxford UP.

Perner, J. (2003), ‘Dual control and the causal theory of action’, in Roessler, J. & Eilan, N. (eds.), *Agency and Self-Awareness*. Oxford: Clarendon Press.

Pollard, B. (2003), 'Can Virtuous Actions Be Both Habitual and Rational?', *Ethical Theory and Moral Practice* 6: 411-425.

Pollard, B. (2005), 'Naturalizing the Space of Reasons', in *International Journal of Philosophical Studies* 13(1): 69-82.

Pollard, B. (2006), 'Actions, Habits, and Constitution', *Ratio* 19: 229-248.

Pollard, B. (2006b), 'Explaining Actions with Habits', *American Philosophical Quarterly* 43: 57-68.

Proust, J. (2003), ‘Perceiving Intentions’, in Roessler, J. and Eilan, N. (eds.), *Agency and Self-Awareness*. Oxford UP.

Reason, J. (1990), *Human Error*. Cambridge UP.

Roessler, J. (2003), ‘Intentional Action and Self-Awareness’, in Roessler, J. and Eilan, N. (eds.), *Agency and Self-Awareness*. Oxford UP.

Rosen, G. (2001), ‘Responsibility and Moral Ignorance’, *NYU Colloquium in Law and Philosophy*.

Rosen, G. (2003), ‘Culpability and Ignorance’, *Proceedings of the Aristotelian Society*, CIII: Part 1, 2003.

Ruben, D.H. (2003), *Action and its explanation.* Oxford UP.

Ryle, G. (1949), *The Concept of Mind*. Penguin.

Sartre, J.-P. (1943), *Being and Nothingness*.

Schlosser, M.E. (2007), ‘Basic deviance reconsidered’, *Analysis* 67 (3): 186-194.

Searle, J. (1983), *Intentionality*. Cambridge UP.

Searle, J. (1990), 'Is the Brain's Mind a Computer Program?', *Scientific American* 262: 26-31.

Searle, J. (1992), *The rediscovery of the mind*. The MIT Press.

Sellars, W. (1956), 'Empiricism and the Philosophy of Mind', *Minnesota Studies in the Philosophy of Science* 1: 127-96.

Shallice, T. (1982), ‘Specific impairments of planning’, *Philosophical Transactions of the Royal Society of London* B298: 199–209.

Shallice, T. (1988), *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.

Shallice, T., & Burgess, P. (1996), ‘The domain of supervisory processes and temporal organisation of behaviour’, *Philosophical Transactions of the Royal Society of London* B351: 1405–1412.

Simester, A. P. (1996), ‘Agency’, *Law and Philosophy* 15: 159-181.

Smith, M. (1987), 'The Humean Theory of Motivation', *Mind* 96: 36-61.

Smith, M. (1996), *The Moral Problem*. Harvard UP.

Smith, M., ‘Cognitivist vs Non-Cognitivist of the Belief-like and Desire-like Features of Evaluative Judgements’, forthcoming.

Stout, R. (1996), *Things that happen because they should.* Oxford UP.

Stout, R. (2005), *Action*. McGill-Queen’s University Press.

Stoutland, F. (1985), 'Davidson on Intentional Behaviour', in LePore, E. and McLaughlin, B.P. (eds.), *Actions and Events*. Basil Blackwell.

Sutton, J. (2007), ‘Batting, Habit, and Memory: the embodied mind and the nature of skill’, in McKenna, J. (ed.), *At the Boundaries of Cricket.* Taylor and Francis.

Sverdlik, S. (1996), ‘Consistency Among Intentions and The ‘Simple View’’, *Canadian Journal of Philosophy* 26: 515-522.

Tanney, J. (1995), 'Why Reasons May Not be Causes', *Mind & Language* 10: 103-126.

Thalberg, I. (1977), *Perception, Emotion, and Action*. Basil Blackwell.

Thalberg, I. (1984), 'Do our intentions cause our intentional actions?', *American Philosophical Quarterly* 21: 249-260.

Thomson, J.J. (1977), *Acts and Other Events*. Cornell University Press.

Van Inwagen, P. (1983), *An Essay on Free Will*. OUP.

Velleman, J. D. (1985), ‘Practical Reflection’, *The Philosophical Review* 94: 33-61.

Velleman, J. D. (1992), 'What Happens When Someone Acts?', *Mind* 101: 461-481.

Vihvelin, K. (2000), 'Freedom, foreknowledge, and the principle of alternate possibilities', *Canadian Journal of Philosophy* 30:1-23.

Vollmer, F. (1993), ‘Intentional Action and Unconscious Reasons’, *Journal for the theory of social behaviour* 23: 315-326.

Wilson, G. (1989), *The Intentionality of Human Action*. Stanford UP.

Withehead, A.N. (1911), *An Introduction to Mathematics*. Holt.

Wittgenstein, L. (1953), *Philosophical Investigations*. Basil Blackwell.

Wittgenstein, L. (1958), *Blue & Brown Books*. Basil Blackwell.

Wittgenstein, L. (1969), *On Certainty*. Basil Blackwell.

Wright, L. (1976), *Teleological Explanations: An Etiological Analysis of Goals and Functions*. University of California Press.